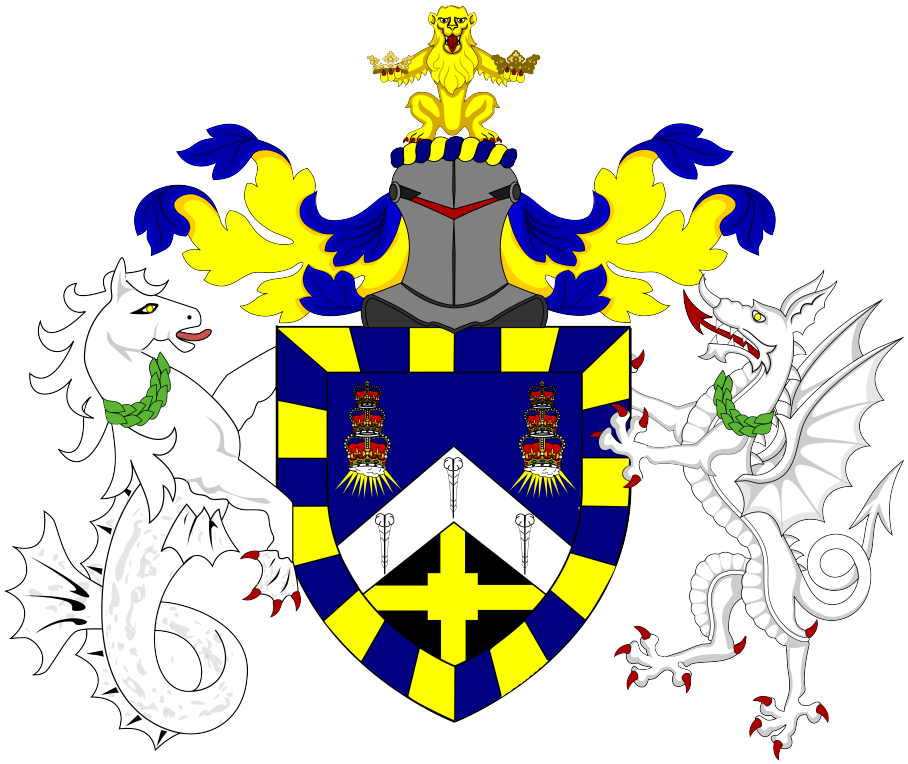
Phosphoproteomic investigation of kinase signalling network plasticity in response to chronic PI3K and mTORC1/2 inhibition

Queen Mary University of London

Barts and the London School of Medicine and Dentistry



Submitted in partial fulfillment of the requirements of the Degree of Doctor of Philosophy

Author: Edmund H. WILKES

Principal Supervisor: Dr Pedro R. CUTILLAS

Additional Supervisors: Prof Bart H. VANHAESEBROECK and Prof Jude FITZGIBBON

Nature uses only the longest threads to weave her patterns, so that each small piece of her fabric reveals the organization of the entire tapestry.

Richard P. FEYNMAN

Acknowledgements

First and foremost I would like to thank Cancer Research UK (CRUK) for providing the financial support that enabled the completion of this PhD. It has been a real privilege to represent CRUK for the duration of this project at institutional, national and international levels. I would also like to thank Barts Cancer Institute (BCI) and Barts & The London School of Medicine & Dentistry for providing the studentship and continued support throughout the project.

This PhD has been a metaphorical and literal journey that has provided me the opportunity to work in many different laboratory environments, learn from many exceptional scientists, and provided me exposure to many diverse areas of science. Firstly, I would like to thank the entirety of the Centre for Cell Signalling (BCI, now University College London) – particularly Mike, Juan-Carlòs, Salma, Pedro, Vinni, Michalina, Inma, Larissa, Roberto, Maria, and Benoit – with whom I spent my first year and who provided me with such an enjoyable introduction to lab-based research. Secondly, I would like to thank members of the MRC Clinical Sciences Centre (Imperial College London) – particularly Anna, Inna, Lucy, and Ari – for welcoming me into their institute and providing the foundation for a very productive second year. I would especially like to thank Pete for his friendship, guidance, and help in editing Chapter 1 of this thesis, and Alex for his mass spectrometric mastery. Thirdly, I would like to thank the Centre for Haemato-Oncology (BCI) – particularly Rob, Tim, Áine, Caroline B., Maria, Arran, Emma, and Ciara – for providing me a supportive and effervescent environment in which to work during the final part of my PhD. I would like to extend these thanks to fellow PhD students at BCI and other institutions – Abbie, Camille, Emma, Laura, Lily, Ketan, Alex, Caroline S., Claire, and many others – who, through so many shared experiences, made this journey so fulfilling and enjoyable. Lastly, I would like to thank my parents, Meg and friends for their endless affection, encouragement, and support in the completion of this degree.

Finally, I would like to sincerely thank my supervisors (Profs Bart Vanhaesebroeck and Jude Fitzgibbon) and tutor (Prof John Marshall) for providing me with invaluable advice and encouragement throughout the project. My greatest thanks however, go to Pedro Cutillas for providing me the opportunity to complete this degree, and for being such an excellent supervisor, teacher, and mentor throughout its duration. It has been a privilege to work as a member of his group and I hope this project will not be the end of the work we do together.

Statement

I, Edmund H. Wilkes, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature: 
Date: 05/02/2015

Details of collaboration and publications at the time of writing:

- Chapter 1; Sections 1.3 and 1.4:

  Wilkes, E. H., Casado. P. and Cutillas P. R. Approaches for measuring signalling plasticity in the context of resistance to targeted cancer therapies, *Biochemical Society Transactions*, 1;42(4):791-7, 2014

- Chapters 4-6:

  Wilkes, E. H., Terfve, C., Saez-Rodríguez, J., and Cutillas P. R. Empirical inference of circuitry and plasticity in a kinase signaling network. *Proceedings of the National Academy of Sciences*, 2015 (under revision)

- Chapter 4:

  Terfve, C., **Wilkes, E. H.**, Cutillas P. R. and Saez-Rodríguez, J. Large scale models of signal propagation in human cells derived from discovery phosphoproteomic data. *Nature Methods*, 2015 (under revision)

Abstract

Class I phosphoinositide 3-kinases (PI3K) and mammalian target of rapamycin complexes 1/2 (mTORC1/2) are enzymes that play important roles in elementary biology and disease. As a consequence, numerous small-molecule inhibitors of their catalytic activity have been developed and these have shown clinical utility in certain cancers. Unfortunately, acquired resistance to these therapies is a common phenomenon and often occurs relatively quickly following treatment.

Our understanding of how resistance develops is hampered by the difficulty of measuring the circuitry and plasticity of the signalling networks that these and other kinases signal within. Advances in mass spectrometric technologies have rendered the routine quantitative interrogation of the phosphoproteome (the set of phosphorylated proteins expressed in a particular biological system at a specific time) more tractable than ever before. The aim of this project therefore, was to improve upon existing mass spectrometry (MS)-based phosphoproteomics methods, and to utilise these to contribute to our understanding of kinase signalling networks and examine their plasticity in models of acquired resistance to PI3K and mTORC1/2-targeted therapies.

Novel approaches for the enrichment of phosphopeptides from complex biological matrices (and their analysis by MS) were designed, tested and optimised. These methods were then used to systematically characterise a kinase signalling network comprising the Akt/PI3K/mTOR and MEK/ERK signalling axes in MCF7 breast cancer cells. The biological relevance of this network was confirmed through the assessment of its dynamics upon EGF and IGF-1 stimulation. Finally, the plasticity of this network following chronic treatment with targeted PI3K and mTORC1/2 inhibitors (GDC-0941 and KU-0063794) was examined in cell-line models of acquired resistance to these two compounds. This revealed that these cells each remodelled this network in a different manner, thus indicating that the initial conditions of the system were not the sole determinant of how resistance was acquired.

Contents

Abbreviations

ACN	Acetonitrile
ALK	Anaplastic lymphoma kinase
AMPK	5'-AMP-activated protein kinase
APEX	Absolute protein expression
AR	Androgen receptor
BSA	Bovine serum albumin
CAMK2	Calcium/calmodulin-dependent protein kinase II
CID	Collision-induced dissociation
CML	Chronic myeloid leukaemia
DC	Direct current
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulphoxide
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediamine tetra-acetic acid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
ER	Oestrogen receptor
ERK	Extracellular-regulated kinase
ERLIC	Electrostatic repulsion-hydrophilic interaction chromatography
ESI	Electrospray ionisation
(F)ETD	(Front-end) electron-transfer dissociation
EThcD	Electron-transfer/higher-energy collision dissociation
FA	Formic acid
FBS	Foetal bovine serum
FDR	False discovery rate
FWHM	Full-width half-maximum
GDP	Guanosine diphosphate
GPCR	G-protein coupled receptor
GPF	Gas-phase fractionation
GTP	Guanosine triphosphate
HCD	Higher-energy collision dissociation
HER2	Human epidermal growth factor receptor 2
HILIC	Hydrophilic interaction chromatography
IAM	Iodoacetamide
ICAT	Isotope-coded affinity tag
IGF	Insulin-like growth factor
IGFR	Insulin-like growth factor receptor
InsR	Insulin receptor
IMAC	Immobilised metal affinity chromatography
INPP4B	Inositol polyphosphate-4-phosphatase
IRS-1	Insulin receptor substrate 1
iTRAQ	Isobaric tag for relative and absolute quantification

KSEA	Kinase-substrate enrichment analysis
LC (nLC)	Liquid chromatography (nanoflow liquid chromatography)
LC-MS	Liquid chromatography-mass spectrometry
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LTQ	Linear trap quadrupole
MALDI	Matrix-assisted laser desorption ionisation
MAPK	Mitogen-activated protein kinase
MAP2K	Mitogen-activated protein kinase kinase
MOAC	Metal oxide affinity chromatography
MS	Mass spectrometry
MSA	Multi-stage activation
mAb	Monoclonal antibody
mSIN	Mammalian stress-activated protein kinase interacting protein 1
mTOR	Mammalian target of rapamycin
NSCLC	Non-small-cell lung cancer
pAb	Polyclonal antibody
PAGE	Polyacrylamide gel electrophoresis
PBS	Dulbecco's phosphate-buffered saline
PDK1	Phosphoinositide-dependent kinase 1
PESCAL	Peak statistic calculator
PI3K	Phosphoinositide 3-kinase
PI	Phosphoinositide
PKC	Protein kinase C
PP2A	Protein phosphatase 2
PRAS40	Proline-rich Akt substrate (40 kDa)
PTEN	Phosphatase and tensin homolog
Q-TOF	Quadrupole time-of-flight
QqQ	Triple quadrupole
RF	Radio frequency
ROCK	Rho-associated protein kinase
RP	Reversed-phase
RPS6KB1	Ribosomal S6 protein kinase B1
RTK	Receptor tyrosine kinase
SCX	Strong-cation exchange
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SEM	Standard error of the mean
SGK3	serum/glucocorticoid-regulated kinase 3
SHIP	SH2 domain-containing inositol 5-phosphatase
SILAC	Stable isotope-labelled amino acids in culture
SIMAC	Sequential elution from IMAC
SRM/MRM	Selected reaction monitoring/multiple reaction monitoring
TBS	Tris-buffered saline
TFA	Trifluoroacetic acid
TMT	Tandem mass tag
TOF	Time-of-flight
TRIS	Tris(hydroxymethyl)aminomethane

TSC	Tuberous sclerosis protein complex
ULK1	Unc-51-like kinase 1
UPLC	Ultra-performance liquid chromatography
XIC	Extracted ion chromatogram

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Phosphoinositide 3-kinases

Phosphoinositide-3-kinase (PI3K) and mammalian target of rapamycin (mTOR) have been identified, amongst other kinases, as having critical importance in both fundamental and disease biology. PI3Ks are a family of enzymes that catalyse the transfer of a phosphate group from the γ position on adenosine triphosphate (ATP) to the 3-hydroxyl groups of phosphatidylinositol lipids at the plasma membrane. They have been shown to be involved in a diverse array of cellular processes including: proliferation, migration, intracellular trafficking and differentiation. The eight currently recognised members of the PI3K family are grouped into three distinct classes according to the phospholipid products that they produce, their pathways of activation and their regulatory subunits: class I (p110α, β, δ and γ), class II (α, β and γ), and class III (Vps34)[1,2]. Table 1.1 provides an overview of the three classes and a brief summary of the distinctions between them (discussed in further detail subsequently). The functions and roles of the class I PI3Ks are the best-characterised of the three classes. Emerging data focused on class II and III, however, are beginning to reveal their impact on normal and disease biology in more detail.

Class	Structural features/functional domains	Catalytic products	Regulators	Activators
I		PI(3)P, PI(3,4)P_2, PI(3,4,5)P_3	p85α/β, p55α/γ, p50, p101, p84, p87	RTKs, Ras, GPCRs
II		PI(3)P, PI(3,4)P_2	-	RTKs, GPCRs(?), Ca^{2+}
III		PI(3)P	Vps15, p150	GPCRs(?) Amino acids(?)

Table 1.1: A summary of the PI3K enzyme family. The PI3K enzyme family consists of three main classes; I, II and III. The differences in catalytic products, regulatory protein partners and modes of activation are highlighted. PI(3)P, phosphatidylinositol-3-phosphate; PI(3,4)P_2, phosphatidylinositol-3,4-bisphosphate; PI(3,4,5)P_3, phosphatidylinositol-3,4,5-trisphosphate.

1.1.1 PI3Ks – class I

The first class of PI3K enzymes is that which has been studied in most depth and consists of four distinct isoforms which are further subdivided into class IA (p110α, β, and γ) and IB (p110δ) catalytic subunits[3,4]. Their mechanisms of activation, regulation and downstream signalling are relatively well-characterised (discussed further in Section 1.1.4, 1.1.5 and 1.1.6). In addition, their substrate specificity has been well-defined, as it has been shown that they selectively phosphorylate PI(4,5)P_2 to produce PI(3,4,5)P_3 *in vivo*[4] (summarised in Fig 1.1). Importantly, their roles in fundamental biological processes (including proliferation, migration, growth, and metabolism) and disease states (e.g. *PIK3CA* [p110α] mutations in various cancers and overgrowth syndromes) have also been elucidated to some extent[5,6].

1.1.2 PI3Ks – class II

As shown in Table 1.1, the second class of PI3K enzymes consists of three members, each representing single catalytic subunits devoid of associated regulatory proteins: C2α, β, and γ[2]. Due to the lack of regulatory subunits, the three enzymes are believed to be regulated, and their

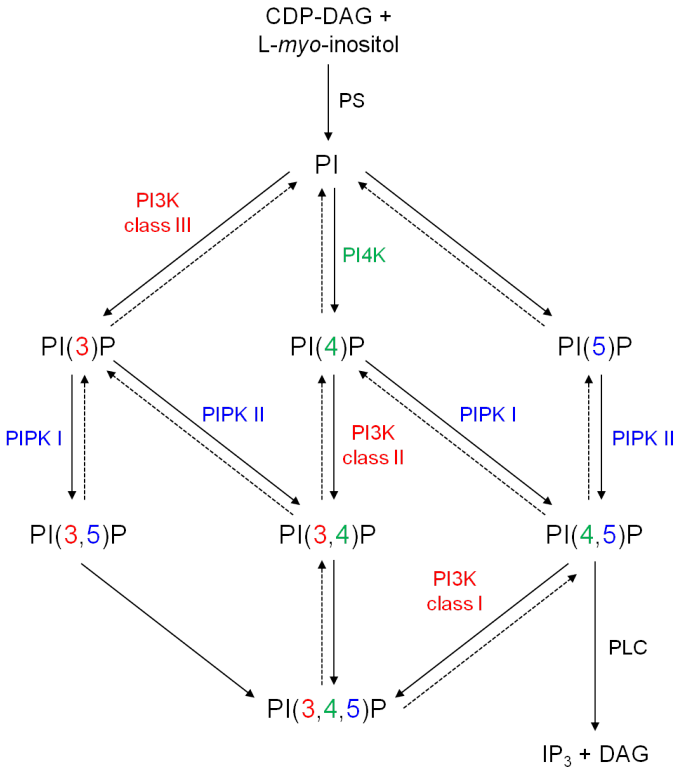


Figure 1.1: An overview of the metabolism of phosphatidylinositols. A host of enzymes catalyse the production and further metabolism of phosphatidylinositols in human cells. Each phosphatidylinositol species possesses an important signalling role, and the equilibria between them is tightly regulated. CDP, cytosine diphosphate; DAG, diacylglycerol; IP_3, inositol phosphate; PI, phosphatidylinositol; PI3K, phosphoinositide 3-kinase; PI4K, phosphoinositide 4-kinase; PIPK I/II, phosphatidylinositol phosphate kinase 1/2; PLC, phospholipase C; PS, phosphatidylinositol synthase.

activity modulated, by their comparatively extended N- and/or C-terminal domains[2]. Evidence suggests that class II PI3Ks interact with a number of cell-surface proteins, namely: clathrin (essential for the endocytosis of multiple cell-surface receptors), epidermal growth factor receptor (EGFR – a cell-surface receptor responsible for activating multiple signalling cascades), and Grb2 (a signalling protein essential for linking Ras signalling to EGFR)[2]. These interactions, amongst other experimental data, point to RTKs and GPCRs being involved in the activation of class II PI3Ks[7].

In contrast to class I PI3Ks, C2α, β, and γ have been shown to possess substrate specificity for PI and PI(4)P, and to subsequently phosphorylate this to produce PI(3)P or PI(3,4)P_2 (Fig 1.1)[8,9]. The majority of the class II-produced PI(3)P is thought to be localised to endosomal compartments where it functions to recruit PX and/or FVYE domain-containing proteins in order to facilitate their downstream signalling events[10].

1.1.3 PI3Ks – class III

The third class of PI3Ks consists of a sole isoform, originally identified in yeast (*S. cevervisiae*), named Vacuolar Protein Sorting 34 (Vps34 – encoded by the *PIK3C3* gene in *H. sapiens*). Vps34 has been shown to utilise PI as its substrate and to catalyse its phosphorylation to form PI(3)P (Fig 1.1). As discussed with regards to class II PI3Ks, the PI(3)P product serves to recruit proteins possessing FYVE and/or PX domains to the intracellular face of the plasma membrane, thus facilitating their roles in downstream signalling pathways and networks[10]. Recent data utilising a selective inhibitor of Vps34 revealed that serum/glucocorticoid regulated kinase 3 (SGK3) is a downstream target of Vps34; this being mediated through its production of PI(3)P and subsequent

binding of SGK3 to these phospholipids via its PX domain[11].

Despite the similarities with class II PI3Ks in terms of the phospholipid products of Vps34's activity, unlike Class II, Vps34 forms a constitutive complex with a regulatory protein, Vps15. Vps15 is a protein kinase that facilitates the activation of the lipid kinase function of Vps34. Vps15 also acts to anchor Vps34 to the plasma and other membranes through its *N*-terminal myristoyl group – thus bringing it into close proximity to its membrane-embedded PI substrates[2,12]. Alongside Vps15's role in the activation and spatial distribution of Vps34, the complex formed between these two proteins nucleates three different multi-protein complexes and the constituents of these act to determine the signalling output produced[2,12]. For example, Vps34 complex I, consisting of Vps34, Vps15, Beclin-1, ATG-14L and NRBF2 is responsible for the regulation of autophagy[13].

The mechanisms of activation of Vps34 still lack clarity; however, amino acids and GPCRs may play a significant role[14,15]. Despite the lack of knowledge regarding the activation of Vps34, several sources of evidence point clearly towards it possessing an important role in biological processes. Firstly, the mammalian and yeast forms of the protein are largely homologous, having been highly conserved through evolution. Secondly, experimental data suggest that it is involved in the regulation of intracellular trafficking and therefore processes such as autophagy, phagocytosis and endocytosis. One recent study determined that AMP-activated protein kinase (AMPK) directly regulates distinct Vps34 complexes – via the phosphorylation of Beclin-1 at Ser^{91} and Ser^{94} – in order to induce autophagy in periods of nutrient stress[16]. Finally, alongside the above, Vps34 homozygous knock-out models in yeast and *D. melanogaster* are not viable, thus highlighting its importance in embryonic development[2].

1.1.4 Class I – mechanisms of activation

Members of PI3K class I are activated by a number of different signalling proteins and their related extracellular factors (summarised in Fig 1.2); the first of which being by stimulation of receptor tyrosine kinases at the plasma membrane (e.g. EGFR and FGFR). RTKs, when stimulated by a multitude of different imflammatory molecules, growth factors, and hormones, are known to activate the p110α, β, and δ isoforms. Stimulation of these receptors by their cognate ligand causes them to dimerise and auto-phosphorylate key tyrosine residues within their intracellular domains, allowing the recruitment of SH2 domains – present within the regulatory subunits of class I PI3Ks – to the receptor. This association induces a conformational change in the dimeric PI3K regulatory-catalytic subunit complex and, as a consequence, the regulatory subunit is no longer able to inhibit the catalytic subunit's activity.

The second mechanism by which a subset of PI3K class I enzymes are activated is via GPCRs. GPCRs, when stimulated to release the $G_{\beta-\gamma}$ subunit by chemokines, cytokines and hormones, have been shown to selectively activate p110γ. Although mechanistically unclear, this mode of activation is believed to involve the interaction of p110γ's regulatory subunits, p101/p84/p87, with $G_{\beta-\gamma}$; thus localising the γ catalytic-regulatory heterodimer at the plasma membrane[17]. Alongside p110γ, recent evidence suggests that p110β is also associated with GPCRs, although this is thought to be regarding its involvement in signalling downstream of the receptors, as opposed to a direct activation mechanism of the kinase[18].

The final class I PI3K activation mechanism, known to activate all isoforms, involves the Ras family of GTPases. Ras interacts with class I PI3Ks via their Ras-binding domains (present in all p110 isoforms – Table 1.1) and a significant body of *in vitro* evidence supports Ras's role in the activation of PI3Ks[2]. Moreover, despite the universality of class I activation by Ras, the combinatorial specificity of Ras and p110 isoforms is functionally relevant; altering the cellular

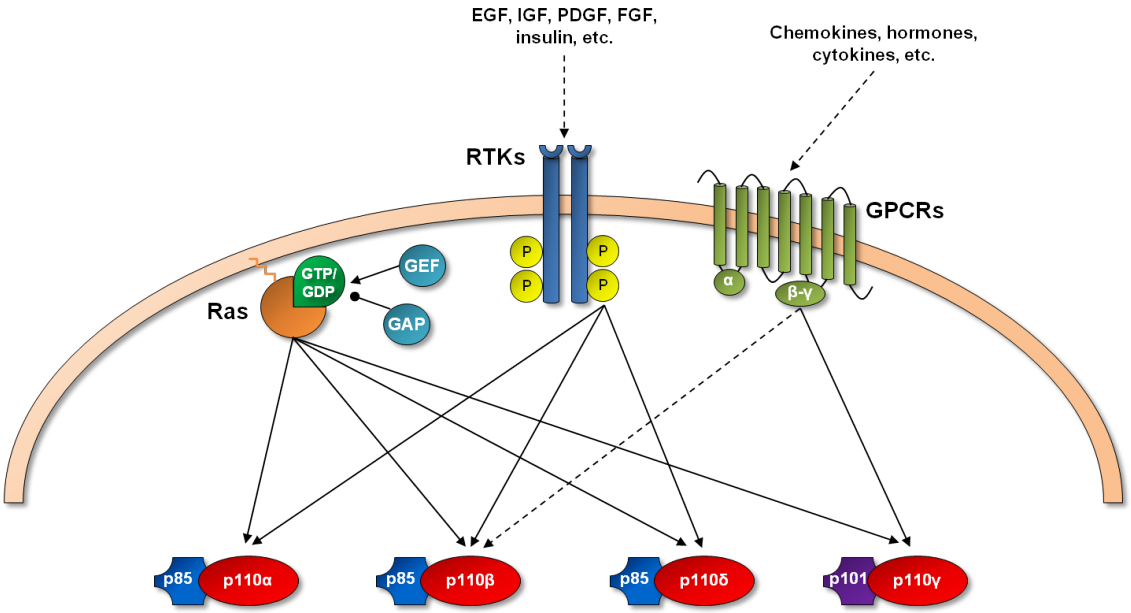


Figure 1.2: Class I PI3Ks are activated by Ras, RTKs and GPCRs. Ras GTPases are involved in the activation of all four class I isoforms, itself being regulated by the action of the relevant GTPase Activating Proteins (GAPs) and Guanine nucleotide Exchange Factors (GEFs). RTKs are responsible for the activation of three of the isoforms when stimulated by growth factors (α, β, and δ). Finally, GPCRs are also involved in the activation of some class I PI3K isoforms; the interaction between $G_{\beta-\gamma}$, p110γ, and potentially p110δ, for exmaple. RTK, Receptor Tyrosine Kinase; GPCR, G-Protein Coupled Receptor; EGF, Epidermal Growth Factor; IGF, Insulin-like Growth Factor; PDGF, Platelet Derived Growth Factor; FGF, Fibroblast Growth Factor

response to stimuli accordingly[19].

1.1.5 Class I – mechanisms of regulation

PI3Ks possess vital roles in the normal homeostasis of cellular proliferation and tissue maintenance. This necessitates tight regulation of the p110 subunits' catalytic functions; ensuring efficient and appropriate growth and proliferative signals to be transmitted. The main mechanism of regulation is via the regulatory subunits' interaction with their respective catalytic subunits (e.g. p85 and p110 – shown to be present in equimolar intracellular concentrations *in vitro*)[20]. Crystal structures of the p85-p110 complex have been solved and demonstrate that p85 interacts with p110 through its C-terminal region (the C-SH2 domain); this binding the N-terminal region of p110[21]. This functions to protect the catalytic subunit from Ras activation and degradation, whilst also inhibiting the catalytic subunit's enzymatic activity in unstimulated cells[22]. Upon stimulation of RTKs, and the subsequent binding of p85 SH2 domains to the key phosphorylated tyrosine residues, this interaction allows p85 to transmit a confirmational change to the p110 subunit, increasing its catalytic activity.

Alongside the direct subunit interactions discussed above, p85 serves as a signal mediator for integrating multiple cellular signals in order to modulate the catalytic subunit's activity. For example, phosphorylation on Tyr^{688} by Src family kinases increases p85-p110 heterodimer activity. Conversely, p85's interaction with the SHP-1 phosphatase has been shown to down-regulate the heterodimer's activity, demonstrating that this regulation is bidirectional[23]. Another example of such an interaction is that recently shown to exist between the tumour suppressor BRD7 and p85α. In this scheme, BRD7 has been shown to enable nuclear translocation of p85α, removing it from the cytosol and therefore preventing its interaction with p110[24]. Interestingly, evidence suggests that the extent of inhibition provided by the p85 regulatory subunits is dependent upon the p85 and p110 isoforms involved. Therefore, the combinatorial flexibility of different isoforms may

be a mechanism of regulation in and of itself[25,26]. In addition to the regulatory role of p85 subunits, class I PI3Ks have also been shown to be regulated by calpains. Recent data has demonstrated that calpains 1 and 2 are responsible for the proteolytic cleavage of the p110α and p85 subunits; reducing the enzyme's catalytic activity and therefore inhibiting PI3K/Akt signalling[27].

Aside from the regulatory impact of the p85 subunits and calpains, another key mechanism of regulation involves several lipid phosphatase enzymes, namely: PTEN (phosphatase and tensin homologue), SHIP1/2 (SH2 domain-containing inositol phosphatase), and INPP4B (inositol polyphosphate-4-phosphatase, type II). These enzymes are responsible for the dephosphorylation of the phospholipids produced both directly and indirectly by class I PI3Ks and therefore provide the antithesis to PI3K activity. A summary of these interactions is shown in Fig 1.3.

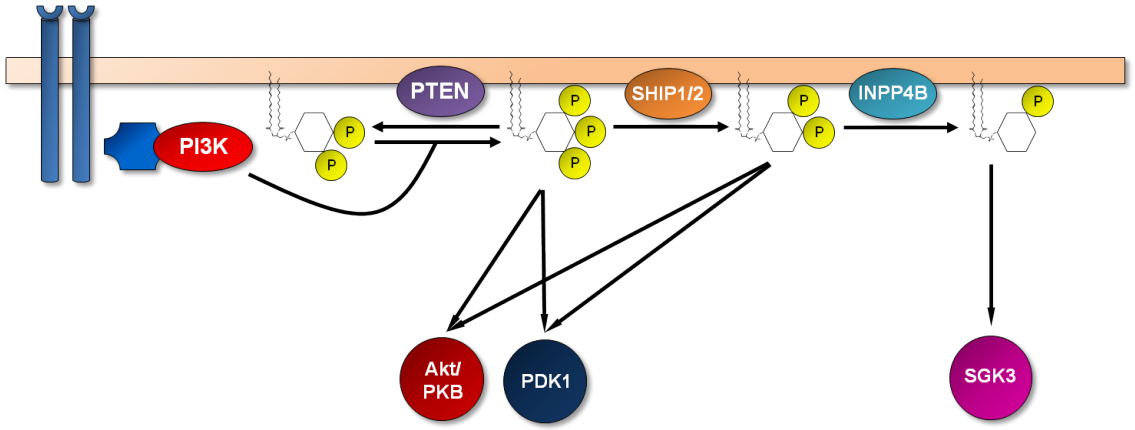


Figure 1.3: Class I PI3Ks are regulated by the phosphatases PTEN, SHIP1/2, and INPP4B. Several lipid phosphatases are known to regulate PI3K activity. This is as a result of their direct dephosphorylation of the phosphoinositols responsible for downstream signalling events (Akt, PDK1, and SGK3 are given as examples). The diagram demonstrates the relationship between RTK stimulation, PI3K activation and the subsequent production of PI(3,4,5)P_3, and how this is then converted by each of the three phosphatases to provide regulation of downstream signalling from PI3K. Akt/PKB, Protein Kinase B; PDK1, 3-Phosphoinositide Dependent protein Kinase-1; SGK3, Serum/Glucocorticoid regulated Kinase-3

PTEN, both a protein and lipid phosphatase, negatively regulates PI3K-based phospholipid signalling by catalysing the dephosphorylation of PI(3,4,5)P_3 to produce PI(4,5)P_2; thus, controlling proliferation, migration and growth signalling[3,28]. It has also been shown that PTEN is negatively regulated by p110δ (signalling through the RhoA/ROCK pathway), giving reciprocity to the relationship between PI3K and PTEN[29]. SHIP1 and 2 also utilise PI$(3,4,5)_3$ as their substrate; however, they catalyse its dephosphorylation on the 5' hydroxyl group to produce PI(3,4)P_2[3]. Both SHIP1 and 2 are known to interact with a considerable number of RTKs at the plasma membrane, alongside the serine/threonine protein phosphatase PP2A, and phosphorylation on key tyrosine residues (and potentially serine/threonine residues – perhaps regulated by the interaction with PP2A) is thought to be vital in their regulation[30]. INPP4B, unlike both PTEN and SHIP1/2, is not involved in the dephosphorylation of PI(3,4,5)P_3 directly produced by class I PI3Ks, but is instead responsible for the removal of the 4' phosphate group from PI(3,4)P_2 to produce PI(3)P. Therefore, as a result of its catalytic activity, INPP4B negatively regulates PI3K/Akt signalling (as both PDK1 and Akt are activated by the PI(3,4)P_2 and PI(3,4,5)P_3 produced by PI3K/SHIP)[31]. Despite this negative impact on Akt signalling downstream of PI3K, INPP4B may be involved in the activation of serum/glucocorticoid regulated kinase 3 (SGK3) – through production of PI(3)P to which SGK3 can bind – thus allowing the propagation of PI3K signalling through Akt-independent routes. INPP4B is found to be selectively expressed in oestrogen receptor (ER)-positive breast cancer cells, and not in ER-negative phenotypes. In addition, INPP4B's role as a tumour suppressor is supported by the strong correlation between INPP4B loss and breast tumours of high

clinical grade and size[31].

1.1.6 Class I – downstream signalling

The phosphatidylinositol lipids produced by PI3K serve to recruit signalling proteins to the plasma membrane and form the major link between membrane signalling proteins (e.g. GPCRs, RTKs, etc.) and cytosolic protein signalling cascades. This recruitment is facilitated by the binding of phosphatidylinositol lipids to pleckstrin homology (PH) domains present in a number of key signalling proteins (e.g. PDK1, Akt, insulin receptor substrate-1 [IRS-1], and the mu isoform of protein kinase C [PKCμ])[32,33]. The crystal structures of several PH domains have been solved, demonstrating the robust presence of two, anti-parallel β sheets, arranged perpendicularly, followed by a C-terminal amphipathic α-helix[34]. PH domains, alongside facilitating interactions with phospholipids, have been shown to bind to certain PKC isoforms and $G_{\beta-\gamma}$ subunits of GPCRs[35,36]. The recruitment of such proteins to the plasma membrane via their PH domains often facilitates their activation and propagation of downstream signalling (Fig 1.3). A classic example of this is Akt, whereby upon binding to $PI(3,4)P_2$/$PI(3,4,5)_3$, it can become phosphorylated by downstream kinases to complete its activation (Thr^{308} by PDK1 and Ser^{473} by mTORC2 – the latter being discussed further in Section 1.2)[36,37].

Alongside the direct facilitation of downstream signalling provided by PI3K through its catalytic activity, it has been proposed that specific isoforms also function as scaffold proteins for multi-protein signalling complexes. For example, evidence suggests that p110γ binds to some PKC isoforms and phosphodiesterase 3B (PDE3B), regulating the respective complexes' functions in a kinase-independent manner[38,39]. Protein-protein interactions of other class I PI3K isoforms are as yet unknown; however, the potential protein kinase activity of PI3Ks suggests that these interactions do exist *in vivo*[3].

PI3Ks' clear impacts as crucial signalling proteins and well-connected nodes of intracellular signalling networks, as evidenced by the rapidly growing body of *in vitro*, *in vivo*, and clinical data supporting this notion, highlights their importance in normal biology whilst concomitantly emphasising their now known role in multiple diseases. PI3Ks' roles in disease are explored further, alongside the closely related mTOR kinase, in Section 1.3.

1.2 Mammalian Target of Rapamycin (mTOR) - function, activation and regulation

mTOR, unlike the PI3K family of enzymes, consists of only one distinct protein kinase molecule; however, it functions to nucleate and provide the catalytic activity for two protein signalling complexes. Each of these complexes possesses a unique set of complex members: mTOR complex 1 (mTORC1; mTOR, mLST8/GβL, Deptor, Raptor, and PRAS40)[40,41] and complex 2 (mTORC2; mTOR, GβL, Rictor, mSIN, and PRR5/PRR5L)[42]. These complexes integrate signalling inputs from a variety of different sources (including growth factors, extracellular amino acids, and insulin) in order to regulate the core, cellular biosynthetic pathways. Despite the similarity in the input signals they receive, their signalling outputs and downstream effectors are quite distinct, but also provide a degree of cross-communication between the two complexes (summarised in Fig 1.4)[40]. Alongside the individual functions of both complexes, the two are deeply embedded in large, multi-protein signalling networks, and cross-talk between them and their downstream effectors has been shown to be key in maintaining cellular and tissue homeostasis.

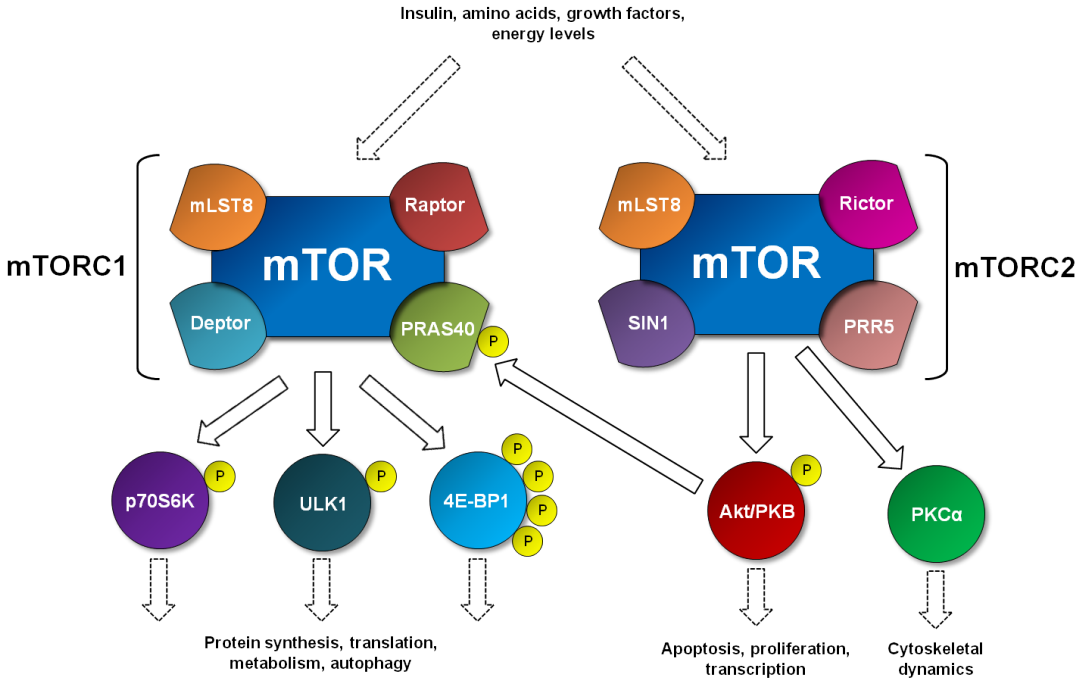


Figure 1.4: The mTOR complexes (1 and 2) and their members. The two mTOR complexes possess both distinct substrates (and therefore biological responsibilities) and complex members; however, significant cross-talk between the two complexes exists. For example, the phosphorylation of PRAS40 (Proline-Rich Akt Substrate 40 kDa) as a part of mTORC1 by Akt, which is itself regulated by mTORC2. mLST8, GTPase β subunit like protein; Deptor, disheveled Egl-10 pleckstrin domain containing mTOR interacting protein; Raptor, regulatory associated protein of mTOR; p70S6K, 70 kDa ribosomal S6 kinase; 4E-BP1, eIF4E-binding protein-1; SIN1, stress-activated protein kinase interacting protein-1; PKC, protein kinase C; Rictor, rapamycin insensitive companion of mTOR.

1.2.1 mTORC1

The first complex mTOR forms is that which has been studied in most depth, primarily due to the discovery that its function is inhibited by the naturally occurring compound rapamycin. mTORC1, as shown in Fig 1.4, is directly responsible for the phosphorylation and activation of several proteins involved in the regulation of cap-dependent translation initiation, namely: eIF4E-binding proteins (the most extensively studied in relation to mTOR being 4E-BP1) and ribosomal S6 kinases (e.g. p70S6Ks)[43]. Through its interaction with the aforementioned substrates, mTORC1 activity is responsible for the stimulation of translation and protein synthesis. One of the mechanisms by which this is mediated is via phosphorylation of 4E-BP1 at Thr^{37} and Thr^{46}; directly leading to 4E-BP1's dissociation from eIF4E. This facilitates the formation of the eIF4F translation initiation complex, and therefore subsequent translation events and protein synthesis[44,45]. The second mechanism involves the mTORC1-regulated phosphorylation of p70S6Ks (RPS6KB1 and 2) at several sites associated with an increase in p70S6Ks' kinase activities (Ser^{394}, Thr^{412}, and Ser^{434} in isoform 1)[43]. Once activated, p70S6Ks are then responsible for the phosphorylation of the S6 protein housed within the 40S ribosomal subunit (Ser^{235}, Thr^{236}, Ser^{240}, and Ser^{244}), eIF4B (Ser^{422}), and programmed cell-death 4 protein (PDCD4 – Ser^{67}), which are believed to regulate translation and protein synthesis[45]. Finally, mTORC1 is also believed to be involved in the indirect activation of RNA polymerases I and III (via MAP1), which are themselves responsible for the transcription of transfer and ribosomal RNAs[46]. Alongside its role in controlling protein synthesis, mTORC1 has also been determined to regulate autophagy. Autophagy is the catabolic process by which cells recycle non-essential cellular machinery in order to survive periods of stress and/or nutrient starvation, and to recycle redundant/malfunctioning cellular components. mTORC1's control of the autophagic process is primarily enacted through its phosphorylation of

ULK1 (Unc-51-Like Kinase 1) at Ser^{757}. This phosphorylation acts to inhibit the interaction of ULK1 with its activatory kinase AMPK (5'-AMP-activated protein kinase) and therefore inhibits overall autophagic flux[47].

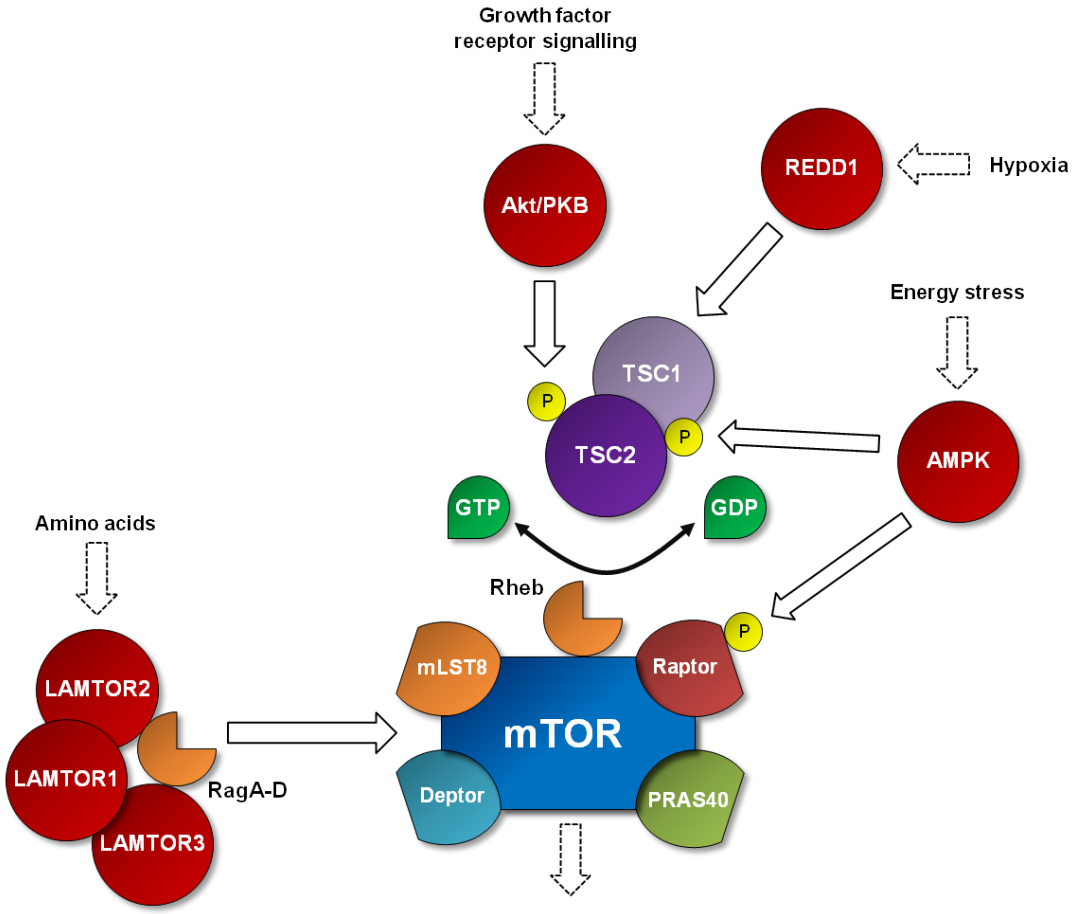


Figure 1.5: The mechanisms of activation and regulation of mTORC1. mTORC1 receives stimulatory and regulatory inputs from four major sources: amino acid levels, growth factor receptor signalling, energy stress, and hypoxia. Regardless of the input, the main regulator of mTORC1 is Rheb, a small GTPase, that when bound to GTP, stimulates mTORC1 activity. TSC1/2, tuberous sclerosis protein 1/2; AMPK, AMP-activated protein kinase; REDD1, regulated in development and DNA damage responses 1; LAMTOR, late endosomal/lysosomal adaptor and MAPK and mTOR activator.

mTORC1 is activated by numerous cellular processes, including: growth factors and hormones (e.g. IGF-1 and insulin), amino acids, energy levels (i.e. ATP/AMP), and hypoxia (Fig 1.5). mTORC1's activity – independent of stimulatory input – is regulated by the small GTPase Rheb (Ras homolog enriched in brain). When Rheb is present in its GTP-bound state, it is reponsible for the activation of mTORC1[48]. Rheb is regulated by numerous proteins; however, the best characterised is the tuberous sclerosis protein (TSC1/2) complex. TSC2 functions as a GAP and thus catalyses the conversion of Rheb-GTP to Rheb-GDP, resulting in a suppression of mTORC1's signalling output[49]. In the context of growth-factor signalling, Akt directly phosphorylates TSC2 (at Ser^{939} and Thr^{1462}), leading to its inactivation, thus providing a direct mechanism by which the PI3K/PDK1/Akt axis activates mTORC1[50]. In addition, Akt also directly phosphorylates PRAS40 (at Thr^{246} – Fig 1.4), which facilitates mTORC1's phosphorylation of PRAS40 at Ser^{183} ([51]). These phosphorylation events result in an increased association of PRAS40 with 14-3-3 proteins and its subsequent degradation, thereby activating mTORC1[52].

The ability of mTORC1 to respond to changes in intracellular energy levels also occurs through the activity of the TSC1/2 complex; however, in this context its activity is regulated by AMPK. As previously mentioned, AMPK, a kinase sensitive to the intracellular ratio of ATP:AMP, is responsi-

ble for the phosphorylation of ULK1 (at Ser^{317} and Ser^{777}), leading to an increase in its activity and therefore autophagy[47,53]. AMPK is also responsible for the phosphorylation of TSC2 at several serine residues (including Ser^{1387}), providing one mechanism by which it functions to suppress mTORC1 activity. Further data suggest however that AMPK also directly phosphorylates Raptor at Ser^{722} and Ser^{792}, which results in an increased association between Raptor and 14-3-3 proteins, which, similarly to PRAS40, result in Raptor's inactivation[54].

mTORC1's amino acid sensitivity is regulated by another protein complex, the Ragulator complex. This consists of three proteins, LAMTOR1-3 (late endosomal/lysosomal adaptor and MAPK and mTOR activator 1), and is reponsible for the recruitment of the Rag GTPases, RagA-D, to lysosomal membranes. As mTORC1 relocates to lysosomal membranes upon the detection of amino acids, it is brought into contact with Rag GTPases by the Ragulator complex, which is then necessary for its activation by Rheb[55,56].

Alongside its previously discussed activation through growth factor receptors, energy levels and amino acids, mTORC1 is also responsive to hypoxia. This responsiveness is derived from mTORC1's interaction with REDD1 (regulated in development and DNA damage responses 1) and AMPK. The expression of REDD1, a HIF-1α-regulated gene, is enhanced in response to hypoxia and DNA damage. The translated protein binds to TSC2, preventing its degradation and therefore inhibiting mTORC1 signalling. REDD1 is itself regulated by GSK-3β (through phosphorylation at $Thr^{23/25}$) and the CUL4A (Cullin 4A)-ROC1 (regulator of cullins 1)-DDB1 (DNA damage-binding protein 1) ubiquitin ligase degradation system[57]. AMPK, activated by hypoxia, regulates mTORC1 through both TSC1/2-dependent and -independent mechanisms, depending on the cellular context. Experiments have demonstrated that in a MEF model, hypoxia-induced AMPK-mTORC1 regulation is TSC2-dependent; however, in a hepatocyte model, this regulation was determined to be TSC2-independent but Raptor and LKB-dependent[58].

1.2.2 mTORC2

It has been observed that not all of mTOR's functions are inhibited when model systems are treated with rapamycin. The discovery of a second mTOR complex opened up the possibility that any rapamycin-insensitive functions of mTOR could be mediated by mTORC2. mTORC2, unlike mTORC1, is not involved in the phosphorylation of substrates whose activities regulate the energy-intensive processes of translation and protein synthesis. Instead, mTORC2 is responsible for the phosphorylation of multiple members of the AGC kinase family (e.g. Akt, SGK1, and PKC)[40,59,60]. The phosphoryation of Akt (at Ser^{473}) by mTORC2 facilitates Akt's complete activation, thus directly linking mTORC2 to the cellular processes downstream of Akt (e.g. cell survival and apoptosis regulation)[59]. Similarly, mTORC2 phosphorylates and activates SGK1 (at Ser^{422}), which is implicated in the regulation of multiple processes such as proliferation, apoptosis and cell volume control[60]. Furthermore, mTORC2's link with PKCα (specifically through Rictor) renders it a key regulator of cytoskeletal dynamics, involving several small GTPases such as Rac1, RhoA, and Cdc42[61].

The mechanisms by which mTORC2 is activated and regulated lack the clarity of those that have been delineated for mTORC1; however, some data have served to elucidate these in recent years. As is the case for mTORC1, evidence suggests that mTORC2 is activated by growth factors in a PI3K-dependent manner; however, the components linking PI3K signalling to mTORC2 remain elusive. One mechanism by which mTORC2 can be activated by growth factor-mediated PI3K signalling that has been delineated however, is via mTORC2's association with the ribosome[62]. Evidence suggests that an increase in insulin/IGF-1-induced PI3K signalling results in an

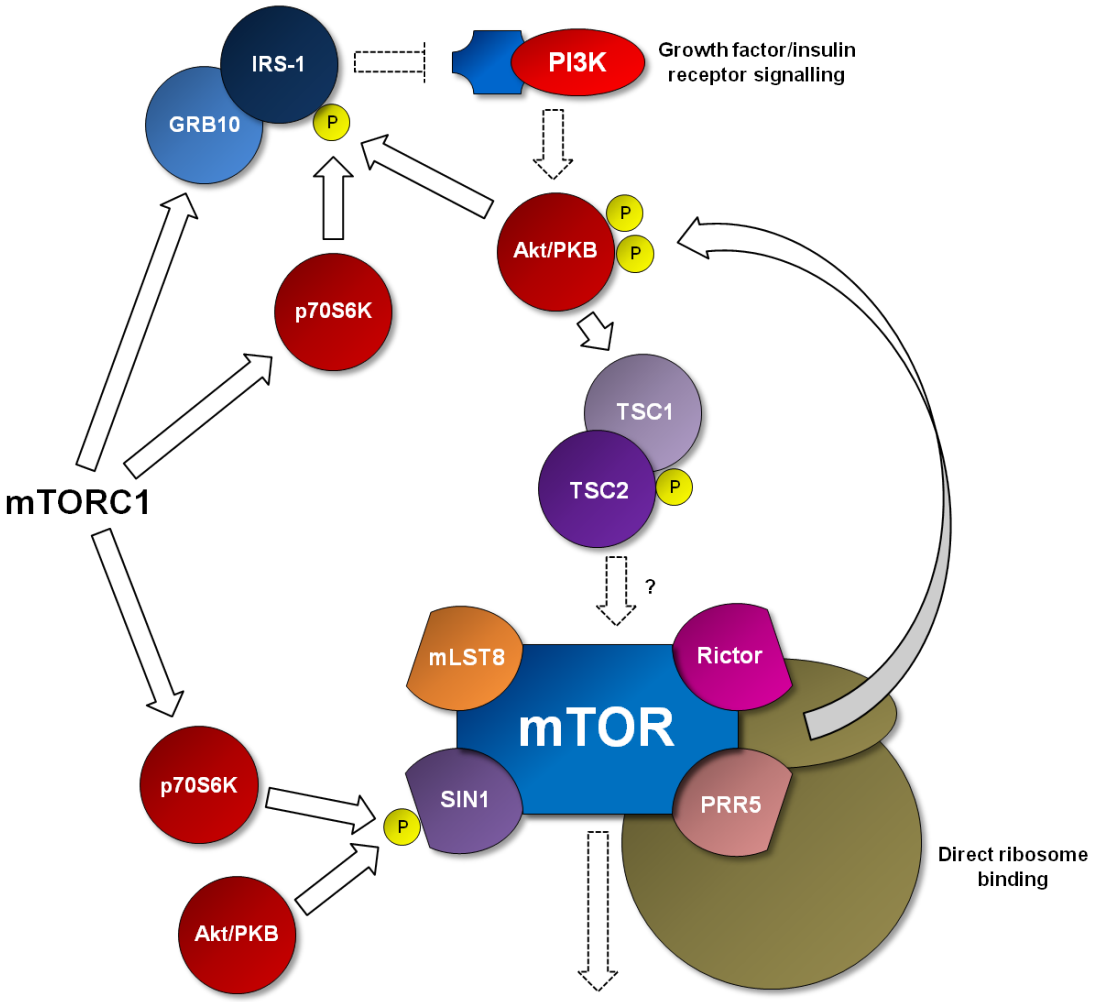


Figure 1.6: The mechanisms of activation and regulation of mTORC2. The mechanisms of activation and regulation of mTORC2 are less clear than those for mTORC1; however, growth factors/insulin receptor signalling and direct ribosome binding are known to cause mTORC2's activation. Unlike mTORC1, mTORC2 is not regulated by Rheb, yet its activation in response to growth factor receptor signalling is TSC1/2-dependent, although the full mechanism of this is not yet understood. There is a large degree of cross-talk between the two mTOR complexes, as evidenced by the negative feedback loop existing from mTORC1's activity through p70S6K and GRB10/IRS-1 working to suppress growth factor-mediated PI3K signalling and p70S6K/Akt's phosphorylation and subsequent destabilisation of mSIN1. IRS-1: insulin receptor substrate-1; GRB10: growth factor receptor-bound protein 10.

increased association of mTORC2 with ribosomes – in a protein synthesis-independent manner – and an increase in its activity as a result (via Akt – Fig 1.6). mTORC2's activity in this context has been demonstrated to be indirectly regulated by mammalian nuclear pre-rRNA processing protein 7 (mNIP7), which is directly involved in the control of ribosome maturation[62]. Interestingly, this activation of mTORC2 via insulin and growth factors has been shown to require the TSC1/2 complex, independent of its GAP activity and therefore independent of Rheb[63,64]; the opposite to mTORC1.

The mechanisms of mTORC2 regulation, similar to those regarding its activation, are much less well-characterised than mTORC1. Despite this, several mechanisms have been elucidated in recent years. Phosphorylation of IRS-1 (insulin receptor substrate-1) and GRB10 (growth factor receptor-bound protein 10) by mTORC1 (and p70S6K) has been shown to provide a negative feedback mechanism to prevent insulin/IGF-1-mediated activation of both mTORC1 and 2 (Fig 1.6)[65]. In addition, recent data has delineated a second mechanism, independent of IRS-1/GRB10, involving mSIN1. In this model, both Akt and p70S6K are responsible for the phosphorylation of mSIN1 at Thr^{86} and Thr^{398}, in a context-dependent manner (Fig 1.6)[66]. These phosphorylation events encourage the dissociation of mSIN1 from mTORC2, thus reducing its activity and signalling output.

1.3 Class I PI3Ks and mTORC1/2 – cancer roles and therapeutic potential

Dysregulation of protein kinase signalling is thought to occur in all cancers. Many examples of such deregulation have now been delineated, including: BRAF activating mutations in melanoma, over-activation of the RTKs epidermal growth factor receptor (EGFR/ERBB1) and human epidermial growth factor receptor 2 (HER2/ERBB2) in lung and breast cancers respectively, and the constituitive activation of the BCR-Abl fusion kinase in leukaemias[67]. Amongst their roles in other diseases, class I PI3Ks and mTOR have also been shown to possess roles in the development of cancers. Firstly, amidst the contributions of all class I PI3K isoforms to carcinogenesis, that of the p110α isoform (*PIK3CA*) is the most well-documented and well-established[6,68–70]. Evidence suggests that there are several routes by which p110α signalling can contribute to tumourigenesis: over-expression of upstream activators, over-expression of p110α itself, activating mutations in the *PIK3CA* gene, and through loss of tumour suppressors (e.g. PTEN loss-of-function mutations)[71]. The involvement of other isoforms is not negligible however, as several studies have reported the differential expression of other class I isoforms in various cancers (p110δ in haematological malignances, for example)[72,73].

Alongside class I PI3Ks, multiple data support the notion that mTORC1 signalling is also frequently aberrant in cancer[40,74]. This dysregulation is thought to be mediated by the loss of key tumour suppressors (e.g. the TSC1/2 complex), which increase mTORC1 activation through PI3K/Akt[75,76]. Experimental data suggest that the phosphorylation states of multiple mTORC1 substrates are dysregulated in cancer (e.g. 4E-BPs and p70S6K) – these correlating with poor prognoses and tumour malignancy[40,74]. The role of mTORC2 in cancer is less well-defined; however, as mTORC2 is potentially activated through EGFR/PI3K signalling, loss of tumour suppressors such as PTEN, could also increase mTORC2's phosphorylation of Akt and other substrates whose downstream cellular processes are implicated in carcinogenesis[76]. The complex relationship that exists between mTOR and PI3K clearly has an important impact on their roles in disease, the extent of the interaction between these kinases and their inter-dependency has yet to be fully explored.

1.3.1 Development of targeted inhibitors against PI3Ks, mTORC1/2 and other kinases

Since the discovery of the involvement of kinases in cancer, malignancies have often been conceptualised as being 'addicted' to the pro-survival signals that they provide when over-activated/ dysregulated. In this, constitutive activation of a pro-survival signalling pathway renders cancer cells reliant on these signals for survival; thus, therapeutic targeting of the over-active pathway inevitably leads to cancer cell death. As a result of these observations, a vast array of antibody-based and small-molecule inhibitors – specific to the deregulated kinase in question – have been developed in recent years. Such inhibitors have become a popular and therapeutically promising area of research; aiming to provide targeted therapy, reduce the side-effects of existing chemotherapies, and overall improve treatment efficacy[77].

Due to the increasing clarity of the roles of mTOR and PI3K in the development of multiple cancers, a selection of the aforementioned small-molecule inhibitors has been developed against their kinase activities. A number have been shown to be successful *in vitro*, in pre-clinical models (e.g. KU-0063794 [dual mTORC1/2][78], Fig 1.7A) and even in clinical trials (e.g. GDC-0941 [pan class I PI3K][79], Fig 1.7B). Despite this, several issues have emerged regarding their and other kinase inhibitors' clinical use, including: inter-patient and inter-/intra-tumour heterogeneity, the choice of target for specific cancers, patient stratification, and tumour resistance[77,80–82]. The latter has become a significant clinical problem as sub-populations of patients fail to respond (despite being predicted to respond based on their genotype), and others that may initially respond well to treatment can relapse, often within a relatively short time-frame[80,81]. As a result, patient survival and quality of life are severely affected. It is important therefore to: (i) understand how only select populations of patients respond to PI3K, mTOR and other kinase-targeted therapies; (ii) predict better those which will benefit from treatment; (iii) elucidate the subsequent mechanisms by which initially responsive cancers are able to evade targeted therapies and become resistant.

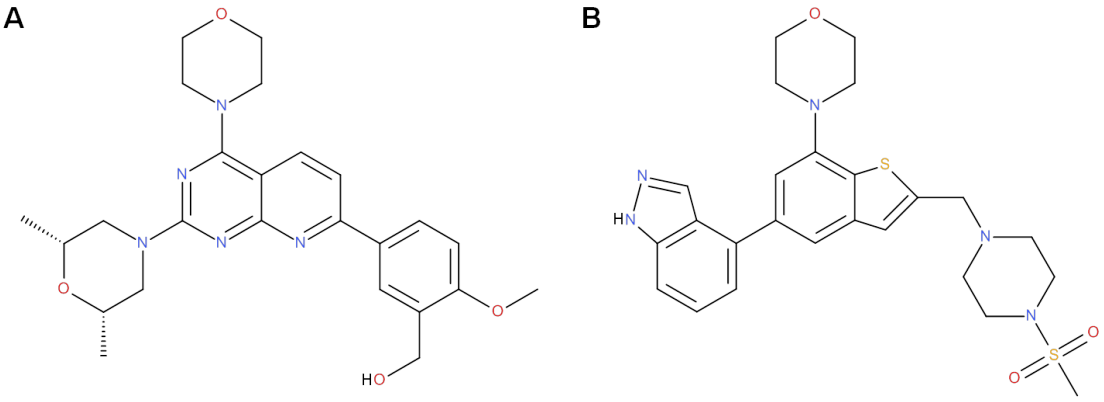


Figure 1.7: The skeletal formulae of KU-0063794 and GDC-0941. (A) rel-5-[2-[(2R,6S)-2,6-dimethyl-4-morpholinyl]-4-(4-morpholinyl)pyrido[2,3-d]pyrimidin-7-yl]-2-methoxy-benzenemethanol: mTORC1/2 inhibitor, KU-0063794. (B) 2-(1H-indazol-4-yl)-6-[[4-(methylsulfonyl)-1-piperazinyl]methyl]-4-(4-morpholinyl)-thieno[3,2-d]pyrimidine: pan-class I PI3K inhibitor, GDC-0941. Both inhibitors are ATP-competitive, catalytic inhibitors, thus binding to the ATP-binding pocket of their respective enzyme.

1.4 Kinase inhibitor resistance

Resistance to kinase inhibitor therapies is generally classifed into two, broad categories: intrinsic (primary) or acquired (secondary). Intrinsic resistance, a lack of response to initial treatment, implies that the cellular populations within a tumour are pre-resistant to targeted therapy. Mechanisms of such resistance have been explored and involve numerous cellular processes, including:

inhibition of negative feedback loops and pre-existing compensatory pathways. In contrast to intrinsic resistance, acquired resistance is the process by which cancerous cells initially respond to therapy, yet develop resistance following prolonged treatment, thus escaping the cytostatic and cytotoxic effects of the inhibition of the targeted kinase. The route by which cancerous cells acquire this resistance is often facilitated by *de novo* genetic changes to the target kinase and/or up-regulation of parallel signalling pathways. Both classes of resistance are discussed in detail in the subsequent sections.

1.4.1 Intrinsic resistance

Despite their success in both pre-clinical and clinical settings, a significant proportion of clinical cohorts do not respond to therapies targeting the kinase that their tumours have been rationalised as being 'addicted' to. As appealing a concept as oncogene addiction is, and despite its canonical success with regards to the treatment of chronic myeloid leukaemia (CML) with imatinib, it can only partially explain the responses to kinase-targeted therapies that are over-active in other cancer types. For example, activating mutations on EGFR are frequent occurrences in non-small-cell lung cancers (NSCLCs). Despite this, clinical trial data demonstrate that the rate of response to an EGFR inhibitor (gefitinib) was only 71.2% in those patients with an activating mutation and 23.5% in those with the wild-type gene[83]. Similarly, clinical data regarding the treatment of V600E mutant BRAF melanomas demonstrated that only 48% of these patients responded to targeted therapy[84]. Moreover, clinical trials of class I PI3K inhibitors demonstrated that only 30% of patients with activating mutations in the *PIK3CA* gene responded to therapy[85]. These examples serve to demonstrate that intrinsic resistance is indeed a significant clinical problem. Furthermore, these data highlight that stratification of patients based upon genetic markers, albeit generally statistically successful, gives rise to a large proportion of false positives (patients that receive treatment but that show no response: 29%, 52% and 70% in the EGFR, BRAF and PI3K studies respectively) and false negatives (patients with wild-type genes that may have benefited from treatment: 23.5% in the EGFR study).

Evidently, the model of oncogene addiction, when invoked in isolation, is not sufficient to account for the observed responses of cancer cells to targeted therapies. Consequently, a more sophisticated model is required in order to rationalise how cancer cells respond to different therapies. A logical assumption is that the wiring of the kinase signalling network in its entirety determines whether a cancer cell population may respond to a given inhibitor. Therefore a concept that is beginning to gain traction is that the activation of both target and parallel pathways contributes to intrinsic resistance. Target and parallel pathways may be activated by feedback loops regulated by the target kinase or may be activated independently. Furthermore, heterogeneous genetic polymorphisms and intrinsically high expression of certain key effector proteins may also allow the continued activation of target and/or parallel pathways and therefore lead to intrinsic resistance.

1.4.1.1 The role of feedback loops in intrinsic resistance

In untransformed cells and tissues, signals are robustly propagated and processed as a result of feedback loops[86]. Emerging data suggest that disruption of negative feedback loops possesses an important role in intrinsic resistance. For example, inhibition of mTORC1 with ATP-competitive inhibitors or with rapamycin and its analogues (rapalogs), results in an alleviation of a negative feedback loop that involves the phosphorylation of IRS-1, an event that decreases the latter's stability. Treatment with mTORC1 inhibitors, in some systems, can therefore lead to sustained

IRS-1 signalling and subsequent activation of PI3K and Akt upstream of mTORC1[87]. A generalised schematic of this is shown in Fig 1.8A. Inhibition of both mTOR complexes prevents this mode of resistance; however, treatment of some breast cancer cells with the mTORC1/2 inhibitor Everolimus has been shown to lead to an S6K-directed feedback loop acting on Ras, PI3K and IRS-1, that results in an up-regulation of MAPK signalling (Fig 1.8B)[88]. The disruption of feedback inhibition has also been observed in HER2 and EGFR-driven cancers treated with a MEK inhibitor. In these systems, MEK inhibition removes the negative feedback from ERK signalling to HER3/ERBB3, thus leading to its hyperactivation and subsequent stimulation of downstream Akt signalling[89]. Therapeutic targeting of Raf has also been observed to cause rapid feedback activation of EGFR[90]. Similar examples have been observed with respect to PI3K inhbiition in PTEN negative prostate cancers, whereby reduced PI3K signalling removes the feedback inhibition from mTOR to the androgen receptor (AR) signalling axis – this being found to lead to a rebound in AR pathway activity mediated by an increase in heterodimeric HER2/3 signalling[91]. Data exploring resistance to the dual PI3K/mTOR inhibitor BEZ235 has shown an increase in signalling from IGF-1R to the JAK2/STAT5 axis, thereby leading to an increase in the production and secretion of IL-8. This secretion of IL-8 was found to provide two functions; increasing CXCR1-JAK2/STAT5 signalling and promoting metastasis[92]. Feedback activation of downstream pathway members has also been shown to occur as a response to HER2 or PI3K blockade, whereby the resultant decrease in Akt signalling leads to an increase in FOXO3a nuclear translocation (this normally being inhibited by Akt), thereby up-regulating the expression of several RTKs (e.g. IGF-1R and HER3/ERBB3) and thus restoring Akt signalling[93,94].

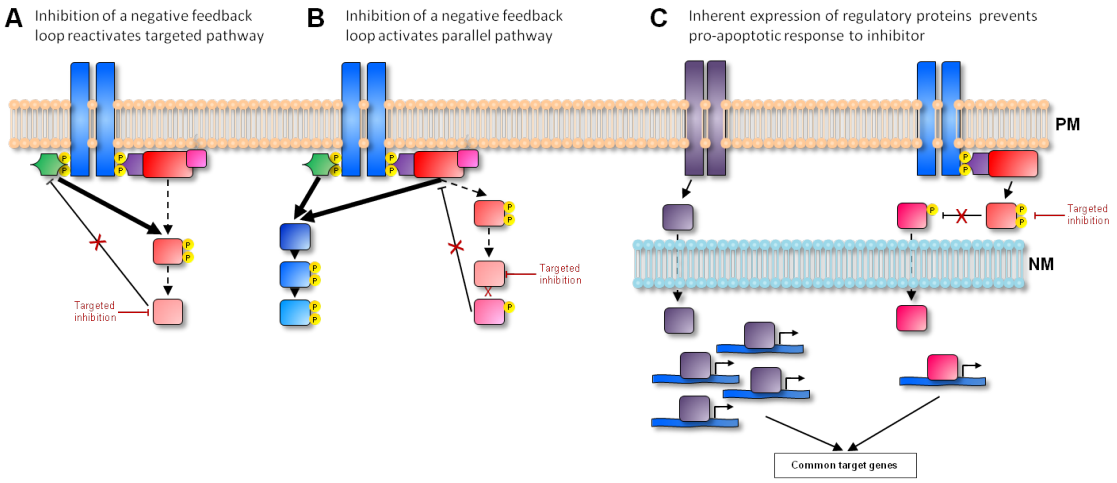


Figure 1.8: Mechanisms of intrinsic resistance to kinase inhibition. (A) Indirect inhibition of an intrinsic negative feedback loop upon targeted inhibition leads to rescued pathway signalling. (B) Indirect inhibition of a negative feedback loop activates a parallel pathway. (C) Inherently high expression of a nuclear protein synergises with the increased nuclear translocation of a transcription factor to maintain a pro-tumourigenic gene expression programme. PM, plasma membrane; NM, nuclear membrane; P, phosphate group. Thick arrows represent increased signalling output; dashed arrows represent indirect activation or multiple signalling steps; solid arrows represent direct activation; flat-headed arrows represent direct inhibition.

1.4.1.2 Parallel factors that contribute to intrinsic resistance

Alongside the disruption of feedback loops associated with resistance, certain tumours also possess resistance due to an intrinsically high activity of certain parallel pathways that can compensate for the inhibition of the target kinase. For example, Akt inhibition with API-2 leads to increased translocation of FOXO3a into the nucleus where it is responsible for the transcription of pro-apoptotic genes (e.g. *BCL2L11*). Intrinsic hyperactivation of wnt-β-catenin signalling (and

therefore high levels of nuclear β-catenin), however, confers the cells resistance to FOXO3a-mediated apoptosis. Once more, the simultaneous nuclear sequestration of FOXO3a and β-catenin appears to lead to the up-regulation of a specific subset of genes that function to promote metastasis (Fig 1.9C)[95]. Further examples also serve to highlight the importance of identified co-activated pathways, as tumours unresponsive to RAF inhibitors demonstrated constitutive Akt activation[96], and AML cells intrinsically resistant to PI3K inhibitors demonstrated concomitant activation of ERK and PKC[97,98]. Adjacent to intrinsic parallel pathway activity, aberrations at the genetic level can also play a role in conferring resistance. For example, recent data have highlighted a common deletion polymorphism in the gene encoding BIM (*BCL2L11*). Due to BIM's pro-apoptotic function, this deletion confers resistance to EGFR therapies in NSCLC and imatinib in CML[99].

1.4.2 Acquired resistance

1.4.2.1 Secondary mutations on the targeted kinase

Cancer cells that do initially respond to targeted kinase inhibition almost inevitably develop resistance after prolonged exposure to the drug. The first mechanism of acquired resistance to be understood regarded mutation of the targeted kinases themselves, preventing the drug from inhibiting pathway signalling (Fig 1.9A and B). Many examples of this phenomenon have now emerged and often involve mutations in the kinase domain that maintain catalytic activity whilst reducing the kinase's affinity for the compound in question[100]. For example, following long-term treatment with Imatinib, CML patients have been shown to often development mutations of a key "gatekeeper" residue near the kinase domain of the over-active Abl kinase (e.g. M351T, T315I, and E255K)[100,101]. This has also been observed in response to Crizotinib (an anaplastic lymphoma kinase [ALK] inhibitor) in NSCLC, whereby patients acquired mutations on ALK (e.g. L1196M and L1152R)[102,103]. EGFR inhibition has also been observed to produce secondary mutations; for example 50% of patients with NSCLC that develop resistance to Erlotinib or Gefitinib presented with the T790M gatekeeper mutation in EGFR[104,105], and colorectal cancers treated with Cetuximab (a monoclonal antibody) presented with mutations in EGFR's extracellular domain (S492R) upon relapse[106].

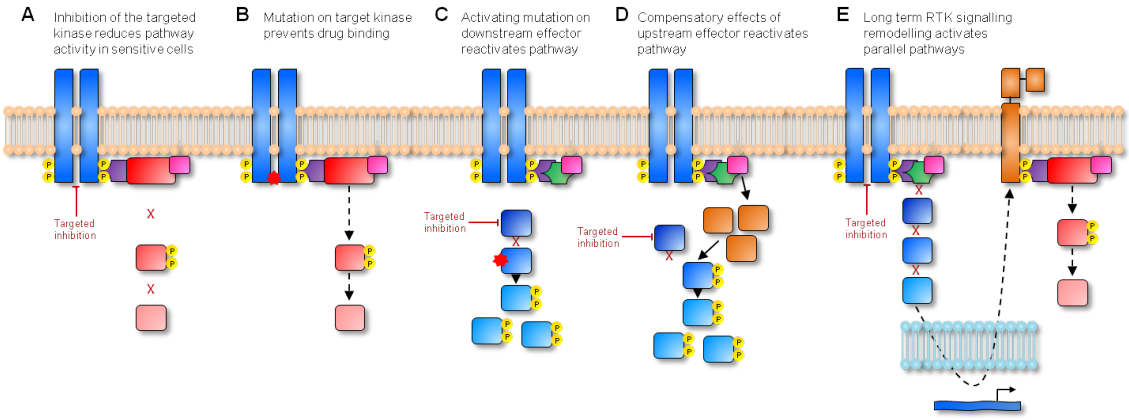


Figure 1.9: Mechanisms of acquired resistance to kinase inhibition. (A) When first treated, the initial inhibition results in reduced pathway signalling. (B) Acquired, secondary mutations on the target kinase abrogates binding of the inhibitor, leading to a rescuing of pathway activity. (C) Secondary mutations on downstream components of the signalling pathway serve to restore activity. (D) Mutation, alteration and/or over-expression of upstream pathway components results in compensatory pathway activation. (E) Kinase inhibition leads to an abrogation of pathway signalling; however, over time the increased expression of compensatory RTKs activates parallel pro-proliferation signalling. P, phosphate group. Thick arrows represent increased signalling output; dashed arrows represent indirect activation or multiple signalling steps; solid arrows represent direct activation; flat-headed arrows represent direct inhibition.

1.4.2.2 Kinase signalling network reprogramming

Several further mechanisms of acquired resistance, independent of mutations on the target kinase, have been described in recent years. Firstly, it has been shown that an increase in the expression/activity of effectors or kinases downstream of the target can compensate for the latter's reduced signalling output. For example, an acquired amplification of eIF4E (at both the genetic and protein level) in response to the PI3K/mTOR inhibitor BEZ-235 has been observed, which provides resistance to this compound[107]. BRAF inhibitor-resistant melanomas have been shown to restore MAPK signalling activity by means of expression of splice variants of the targeted kinase (e.g. p61-V600E-BRAF), mutation of kinases downstream of BRAF (e.g. C121S and Q90P on MEK – Fig 1.9C), and over-expression, mutation, or substitution of upstream pathway activators (e.g. COT, CRAF, NRAS and EGFR/SOS – Fig 1.9D)[108–112].

Cancers are also able to acquire resistance to a targeted compound through the activation of network branches and parallel pathways that are independent (canonically) of the targeted pathway. One such mechanism that has been observed on numerous occasions is the up-regulation of RTKs that then activate compensatory pathways and thus facilitate circumvention of the anti-proliferative effects of the drug. For example, Gefitinib-based anti-EGFR treatment has been shown to lead to the amplification of Met (an RTK) which subsequently activates PI3K/Akt signalling via HER3/ERBB3 (Fig 1.9E)[113]. c-Met amplification has also been observed with respects to resistance to anti-VEGF therapy in glioblastoma[114]. Other data have shown that HER2/ERBB2 is up-regulated in response to Cetuximab-based anti-EGFR therapy; this being mutually exclusive to the T790M secondary mutation on EGFR[115]. Similarly, numerous data have demonstrated that several RTKs are up-regulated in response to BRAF inhibitor therapy in melanoma, namely: EGFR, IGF-1R, and PDGFRβ[110,116,117]. Further studies have also shown the acquired activation of intracellular kinases in response to therapy. For example, PTEN and RB1 deletion/inactivation in BRAF-inhibitor treated melanoma, therefore bypassing ERK signalling reliance through the activation of Akt. Such phenomena have also been observed in the context of long-term PI3K inhibition involving several kinases and pathways such as: MEK in murine lung cancers[118] and PAK1 in lymphoma cell-lines (being activated by PI(4,5)P_2 pools at the plasma membrane)[119].

1.4.3 Understanding resistance mechanisms and kinase signalling at a systems level

A concerted research effort has revealed, as discussed above, that there is a plethora of mechanisms by which cancer cells can intrinsically circumvent, or indeed evolve resistance to, the anti-proliferative/pro-apoptotic effects of a particular kinase inhibitor. This renders the prediction of cancer patients' responses to different therapies a serious challenge. These studies have also highlighted the substantial and functionally important cross-talk that exists between signalling pathways that have previously been conceptualised as being secular and linear[120–122]. Consequently, it has become apparent that in order to fully understand the way in which kinases signal at a fundamental level – and indeed how this signalling is remodelled in disease states – the networks that these kinases form must be studied as a whole[123,124].

Analysis of kinase signalling networks – products of millions of years of evolution – has revealed that they are highly robust and plastic structures[125]. As such networks are responsible for the integration and processing of signals, one could argue that the translation of the concepts of kinase inhibitor resistance over to the robust clinical analysis of patient samples would require a method (or methods) that allow(s) the profiling of the topology and activity of kinase signalling networks in their entirety, without bias. The ultimate aim of such a method would be to transition the field towards the measurement of individuals' malignancies and to facilitate successful ther-

apeutic decisions therefrom. Amongst several methods being used to tackle these challenges, those based on mass spectrometry have proven invaluable in the provision of a more sophisticated understanding of kinase signalling. This is primarily due to their high-content nature, which has allowed a departure from pre-conceptions of which kinases and signalling pathways may be important in a particular system. This provides a move away from inherent bias concerning particularly well-characterised signalling nodes. The impact of MS-based proteomics methods in this context is discussed further in the remainder of this chapter.

1.5 Proteomics

1.5.1 Mass spectrometry – applications to proteomics

Proteomics is the study of the entire complement of proteins within a biological system at a specific point in time or under a specific, controlled condition. Amongst other techniques that have been applied to proteomics, mass spectrometry is one that has revolutionised the field. At its most fundamental level, mass spectrometry is the measurement of the mass-to-charge ratio (*m/z*) of gas-phase ions. From this conceptually simple information it is possible to determine, when coupled to chromatographic techniques, the composition of complex mixtures of compounds. Due to its inherent specificity, capacity and unbiased nature, it has relevance in numerous areas of research and industry and is used for the analysis of a wide variety of chemical and biochemical moieties.

Despite some application-specific variations, all mass spectrometers are comprised of the same, fundamental units: an ion source, a mass analyser, and a detector. As mentioned above, the analyte of interest must be in the gas-phase in order to be analysed by the mass spectrometer. The first element of a mass spectrometer, the ion source, is responsible for the introduction of gas-phase ions into the instrument – this process often involving the transfer of the analyte from solid or liquid phase to the gas phase. Ions are then introduced into the mass spectrometer through the use of a voltage and pressure gradient (e.g. a large negative voltage and pressure being applied to the ion transfer capillary when positive ions are being analysed). Once in the mass spectrometer, ions are separated according to their specific *m/z* in the mass analyser; this step being of great importance, as it allows the distinction of the different molecular species in the sample. The ions then pass on to the detector where their relative abundance in the sample can be measured by monitoring the charge or current produced as each ion hits the detector. Plotting the relative abundance of the ions detected in the mass spectrometer versus their unique *m/z* produces a mass spectrum. Once constructed, mass spectra can then subsequently be used to calculate the charge (*z*) of each ion of interest and thus the absolute mass (*m*) of each ion can be determined.

Mass spectra produced from the process outlined above contain both qualitative (the unique *m/z* peaks present) and quantitative (relative peak abundances) information, and are thus invaluable in the study of chemical and biochemical mixtures. As a result, the application of mass spectrometry to the field of proteomics has flourished over the last decade; particularly as the resolution, sensitivity, speed, and availability of instruments has greatly increased[126]. The use of mass spectrometry in proteomics has largely, but not exclusively, been involved in the analysis of purified peptide mixtures (a so-called "bottom-up" approach)[127]. This is because – alongside the previously mentioned quantitative capabilities of mass spectrometers – the introduction of tandem-mass spectrometry (MS/MS or MS^2) and the polymeric nature of peptide structures have made it possible to sequence peptides in an unbiased and relatively high-throughput man-

ner. Consequently, it became possible to analyse the proteome of biological systems without prior knowledge of their protein composition. This not only allowed the sequencing of novel protein variants, but also, with the advent of additional tandem-MS fragmentation methods (e.g. MS^3 and collision-induced dissociation–multi-stage activation [CID-MSA]), the robust identification of multiple known and novel post-translational protein modifications[128]. Taken together, mass spectrometry-based proteomics overcomes many of the limitations present in genomics or transcriptomics analyses, and thus presents itself as a powerful tool for the study of complex biological systems.

1.5.2 Mass spectrometers for proteomics

Before its application to proteomics, mass spectrometry was used (and remains popular) for the analysis of small molecules. This generally involves the separation of analytes via gas chromatography, whereby the molecules of interest are passed through and separated on a column in the gas phase by exchanging with a stationary phase according to their specific physico-chemical properties. Gas chromatography is not well-suited to proteomics analyses however, as peptides are relatively non-volatile. As a result, the development of ionisation techniques that could produce gas phase ions from aqueous, liquid phase samples was a major boon for the proteomics field. Electrospray ionisation (ESI) is such a method, and is that which is most often used in proteomics studies. ESI, summarised in Fig 1.10, involves the application of a high voltage to the analytes as they pass through an electrically conductive capillary in the liquid-phase. As a result, an aerosol of multiply charged solvent droplets forms[129]. Two models have been developed to describe the process by which individual ions are ejected from the solvent droplets that form: the ion evaporation and charge residue models. Both models end with the production of desolvated, gas-phase ions by Coulombic repulsion; however, both propose slightly different parameters which influence the rate of solvent evaporation[130–132]. Nevertheless, the gas-phase ions that are produced are then encouraged to enter the mass spectrometer via a voltage and pressure gradient imposed on the ion transfer capillary (Fig 1.10).

Several variants of standard ESI exist; however, perhaps the most important with respects to proteomics analyses is nano-ESI (nESI). This technique involves identical principles to that of standard ESI, but is performed at much lower flow rates (typically 0.3-0.8 $\mu L \cdot min^{-1}$). nESI has been shown to greatly increase the sensitivity of LC-ESI-MS analyses and reduce ion suppression artefacts[133,134]. As a result, it is now commonplace in proteomics experiments.

Another popular ionisation method utilised for proteomics is Matrix-Assisted Laser Desorption Ionisation (MALDI). MALDI requires the sample to be embedded in a solid-phase matrix (commonly 2,5-dihydroxybenzoic acid – DHB), which is then vapourised by a laser beam to produce protonated (and therefore charged) matrix molecules in the gas phase. The matrix molecules then transfer this charge to the analyte molecules, producing gas-phase, charged analyte ions which can then be analysed in the mass spectrometer[135]. Unlike ESI, MALDI is not easily coupled to liquid-chromatography, and the matrix:analyte ratio – which is often highly variable – has severe impacts on the reproducibility of the method[136]. Furthermore, although both methods are "soft" ionisation techniques, ESI produces more multiply charged ions (typically 2^+ or 3^+ for tryptic peptides), which are better suited for subsequent MS/MS fragmentation and therefore peptide identification.

A large variety of mass analysers are utilised for the study of peptides and proteins, namely: time-of-flight (TOF), quadrupole (Q), linear ion-trap (LIT), orbitrap, and Fourier-transform ion cyclotron resonance (FT-ICR) instruments. TOF instruments rely on the principle that when ions

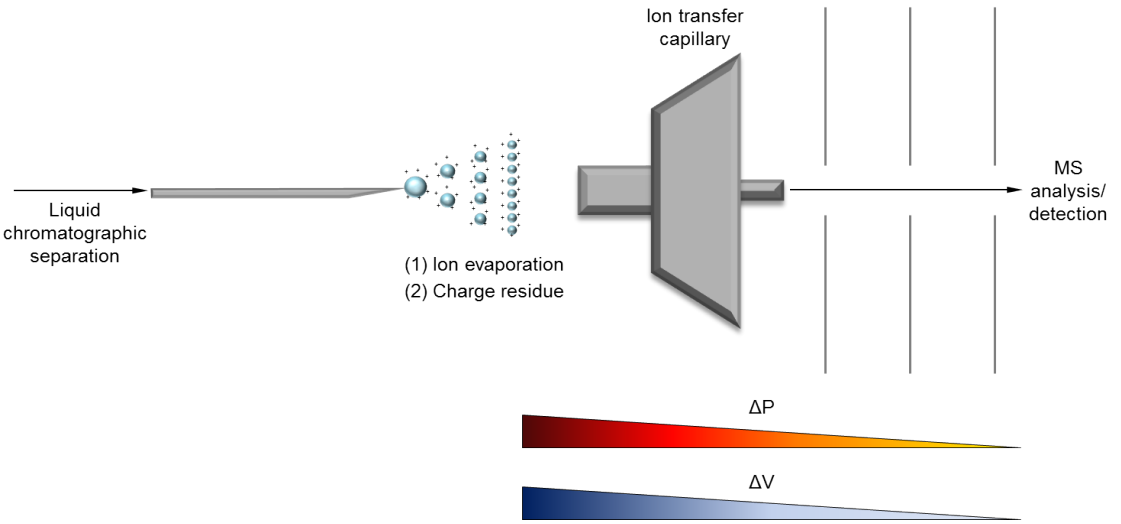


Figure 1.10: The electrospray (ESI) process. Peptide or other analytes are directly infused into the MS or undergo chromatographic separation. The solution then passes through a silica or metallic emitter which has a large voltage applied to it. Charged droplets begin to form on the tip of the emitter; gradually decreasing in size as solvent evaporates. Eventually, single, desolvated, gas-phase analyte ions are formed, and they enter the mass spectrometer via the ion transfer capillary down a pressure (red) and voltage (blue) gradient. The ions are then available for mass analysis and detection.

are accelerated by an electric field and then allowed to drift through a vacuum, they will separate/resolve according to differences in their *m/z*. The flight time (t) of an ion is related to the square-root of the *m/z* of the ion multiplied by a proportionality constant defined by particular instrument characteristics (k). Due to this known relationship, it is possible to calculate the *m/z* of any ion given its measured time-of-flight[137].

Quadrupole mass analysers consist of, at their most fundamental level, four, symmetrically-spaced, parallel rods. They rely on the application of radio frequency (RF) voltages between each pair of rods and subsequent superimposition of a direct current (DC) voltage on the rods. At any given combination of voltages that are applied, only ions of a particular *m/z* will be able to travel with a stable trajectory to reach the detector, and thus, by manipulating these voltages, ions can be separated based upon their unique *m/z*.

Linear and 3D ion-traps are so-called because they are able to trap ions in space through the application of oscillating alternating current (AC) and DC electric fields. The trapped ions can then be ejected from the trap in a *m/z*-specific manner. Furthermore, linear ion traps are able to operate both as a mass-selective filter (i.e. like a quadrupole) or as an actual trap. FT-ICR instruments trap ions using a magnetic field (unlike a linear or 3D ion trap), and then utilise an electric field orthogonal to the magnetic field to excite the ions. This excitation causes the ions to rotate at their specific cyclotron frequency, which is directly related to their m, z, and the magnetic field strength (B). Once excited, the rotating ions induce a charge on the detection electrodes, which is subsequently detected as an image current; critically this can then be converted by a Fourier transform to produce a mass spectrum. The operating principle of orbitraps is very similar; however, they use an electric field to trap the ions, which, when injected into the orbitrap, begin to oscillate around the central electrode at a frequency directly related to their *m/z*[138]. Much like FT-ICR instruments, the image current induced by the oscillating ions passing the detection electrodes is converted by Fourier transform to produce a mass spectrum. Due to their high mass accuracy (1-2 ppm), high resolution (as high as 480,000 FWHM), amenability to hybridisation with other MS analysers, high sensitivity (especially given technological advancements in the ion traps and optics preceding the orbitraps themselves), and relatively high dynamic range, Orbitraps have

become a staple of proteomics analyses in recent years.

1.5.3 Experimental strategies and sample preparation in proteomics

Correct sample preparation is key to the success of any proteomics experiment. There are three conceptual methods by which proteomics experiments can be performed and the samples prepared: bottom-up, middle-down, and top-down[127,139]. The large majority of proteomics experiments to date have employed a bottom-up strategy, whereby the protein mixture of interest is enzymatically digested using a protease of known specificity to yield a complex peptide mixture (each peptide usually being approximately 8-25 residues long, depending on the protease used to create them). The abundance and identity of the analysed peptides are then extrapolated to make inferences as to the quantitative and qualitative properties of the proteins within the sample.

Numerous methods by which proteins can be enzymatically digested have been developed. These methods employ several different proteases, each with their own complement of advantages and disadvantages. Despite this, all the aforementioned methods fall into two main groups: immobilised or in-solution digests. A key consideration of both of these is the way in which interfering substances in popular lysis buffers used for protein analysis can be removed or their action abrogated prior to digestion.

Immobilised digestions rely on the immobilisation of proteins on a solid support in order to allow the liquid-phase removal of contaminants and inferring detergents or chaotropes, and increase the efficiency of digestion. Two classic examples of this are: in-gel digests (where the proteins are embedded in a polyacrylamide gel) and filter-assisted sample preparation (FASP – where the proteins are immobilised on a solid filter support)[140].

In-solution digestions instead rely on the dilution of interfering chaotropes (e.g. urea or guanidine hydrochloride) prior to digestion, so as to abrogate their impact on the enzyme being used. The disadvantage of these methods is that the enzyme remains in the sample and thus slightly increases the peptide complexity of the sample. To counteract this, in-solution reactions can be carried out with an immobilised-bead form of the enzyme of choice, which can be removed by centrifugation at the end of the reaction[141].

Alongside whether the digestion is performed with immobilised or in-solution proteins, another vital consideration is the proteolytic enzyme being used. By far the most popular enzyme used to date is trypsin, a pancreatic serine protease which selectively cleaves peptides at the carboxyl (C) terminus of either arginine (R) or lysine (K) residues[142]. Despite its clear utility, trypsin is not the optimum enzyme to use in all experiments; particularly when the post-translational modifications of interest lie on R/K residues, or when the protein of interest has very few R/K residues (approximately 10% of the theoretical proteome)[143]. As a result, other enzymes have been used, for example: Arg-C, Lys-C, chymotrypsin, pepsin, and Glu-C[139,144]. Recent data has highlighted that the results obtained from a proteomics experiment are highly dependent on the protease used[145]. Moreover, it has become increasingly appreciated that, due to the complementarity of the information gained with the available proteases, a combination is required to reach the required proteome depth[143].

Top-and middle-down proteomics experiments differ from bottom-up strategies in that they involve the analysis of whole proteins or very large peptides respectively. Top-down approaches, often analysing intact, folded proteins, can be used to obtain 3D structural information and the binding affinities of the protein of interest with small molecules or other proteins – this being made particularly accessible in recent years with the mainstream availability of ion-mobility cells and hydrogen-deuterium exchange (HDX) platforms[146–148]. Middle-down methods employ an enzy-

matic digestion step; however, the peptides produced are usually much bigger than those in a bottom-up strategy (2-20 kDa)[144]. This can be particularly useful in distinguishing isoform variants and the study of combinatorial post-translational modifications. Despite their advantages, the complexity of whole-protein and large peptide fragmentation patterns can be a hindrance in the analysis of very complex biological samples. Recent developments in whole protein fragmentation and ion excitation methods (e.g. front-end ETD [FETD] and ion-ion proton transfer [IIPT]), however, have drastically improved the sequencing of proteins in a top-down manner[149].

1.5.4 Liquid chromatography

Due to their inherent complexity, protein or peptide mixtures must be separated before they enter the mass spectrometer for analysis. This is due to the fact that the dynamic range and resolution of mass spectrometers are limited, and thus distinguishing between both high and low abundance species and peptides of similar mass in a single spectrum would be extremely difficult, or indeed impossible. Such separations can be done "on-line" and "off-line", each with their strengths, limitations, and appropriate uses. A common and powerful separation method is liquid chromatography (LC), which is amenable to both on-line and off-line use.

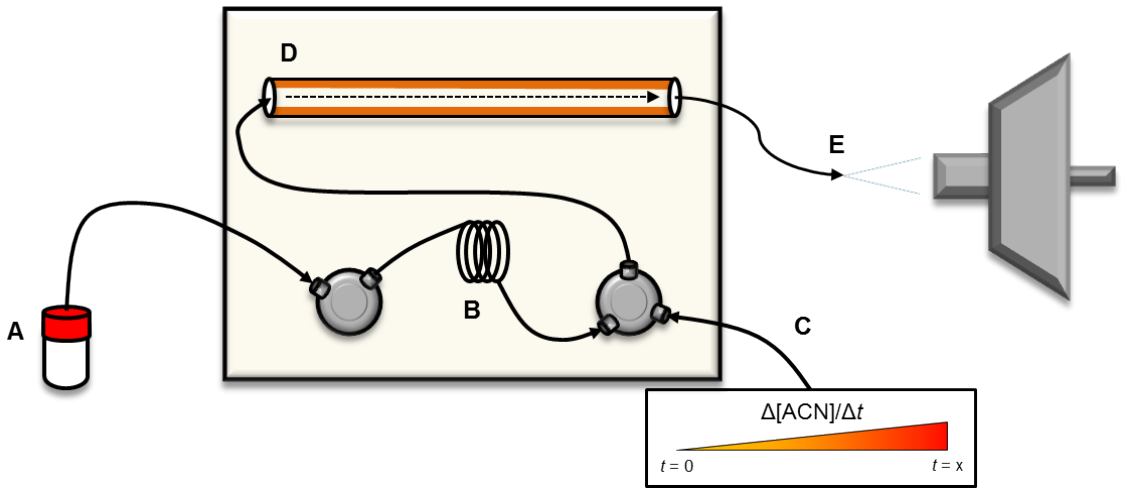


Figure 1.11: The overall principles of reversed-phase liquid chromatography (not to scale). The sample (A) is injected into the sample loop (B). The sample is then loaded onto the column under aqueous conditions (C: t = 0). Peptides bind to the stationary phase (D [orange] – all peptides will bind under aqueous conditions), and are eluted-off at different time-points along the gradient of increasing ACN (C), their individual retention times being a function of their hydrophobicity. As the peptides elute off the column they are directed to the electrospray source (E) and enter the mass spectrometer (or other detector) for analysis.

Liquid chromatography is the process of separating molecular species according to their differential physico-chemical properties in liquid phase. A number of peptide properties can be exploited for this purpose: e.g. pK_A, hydrophobicity (ϕ), and charge, for example. The most commonly exploited peptide property in proteomics – by means of reversed-phase chromatography – is hydrophobicity. The general principle is that when peptides are loaded onto a column containing a hydrophobic stationary phase under aqueous conditions, they will bind to it with different affinities according to their overall hydrophobicity. When a linearly increasing gradient of organic solvent (commonly acetonitrile [ACN/MeCN]) is then passed over the stationary phase, peptides elute along the gradient as a function of their hydrophobicity; the less hydrophobic peptides eluting earlier in the gradient, and *vice versa* (Fig 1.11). Each peptide is then assigned a retention time (t_R) depending on when it elutes from the stationary phase and is detected in the mass spectrometer (or an alternate detector). This chromatographic separation results in a much less complex mixture of peptides entering the mass spectrometer at any given time, and thus increases the

sensitivity, dynamic range, and therefore the amount of information that can be determined from the sample.

Alongside the on-line approaches outlined above, a number of orthogonal, off-line pre-fractionation/separation methods have been used to reduce sample complexity prior to on-line LC-MS analysis. Such methods include 1D or 2D chromatographic or electrophoretic separations, or the use of affinity purification (e.g. metal oxide or antibody) to enrich the sample for proteins/peptides of particular interest. Recent studies have shown that extensive pre-fractionation of particular proteomes with 2D chromatographic methods (commonly strong cation exchange [SCX] and hydrophilic interaction chromatography [HILIC]) can drastically improve the depth of analysis[150]. This, however, comes at the cost of greatly increased sample preparation and MS analysis time, and a substantial increase in the complexity of downstream data analysis. Moreover, a certain amount of sample loss is inherent in such a lengthy method, thus increasing the amount of starting material required.

1.5.5 MS analysis – qualitative proteomics

The power of MS when applied to proteomics is its ability to be used for both targeted and untargeted global analyses. Targeted analysis of proteins and peptides permits their now routine identification; however, current state-of-the-art technology allows for the identification of tens of thousands of peptides and thousands of the corresponding proteins in a comparatively small time-frame, in a unbiased manner. As briefly mentioned in Section 1.5.2, this is mainly due to the introduction of tandem mass spectrometry (MS/MS or MS^2), whereby the mass analysis of different populations of ions is separated. The mass analysis of the ions can be either separated in space (i.e. the mass analysers are physically distinct, although connected) or in time (the ions are measured, separated, and then measured again in the same analyser at $t + x$)[151]. Fig 1.12 provides an overview of the principle of tandem MS in space.

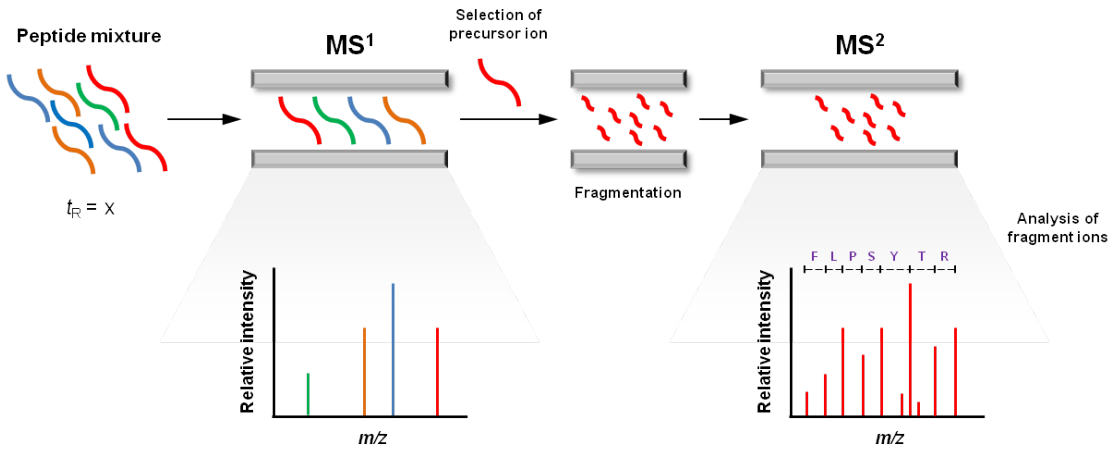


Figure 1.12: The principles of tandem MS. Peptides eluting from the LC system at a particular retention time (x) are introduced to the mass spectrometer and their *m/z* values are measured in the first mass analyser (MS^1). A particular precursor ion (either targeted or non-targeted) is selected and fragmented to yield its constituent fragment ions (a mixture of different sequences that make-up the original peptide). The *m/z* values of these fragment ions are then measured in the second mass analyser (MS^2), and using these data, the sequence of the precursor peptide ion can be determined; either through comparison to peptide databases or *de novo* from manual interpretation of the spectrum.

Fragmentation of ions between MS scans revolutionised the study of proteins, peptides, and their post-translational modifications. Two main varieties of ion fragmentation exist: in-source and post-source. In-source fragmentation describes the process by which ions are given sufficient internal vibrational energy during ionisation to fragment as they enter the mass spectrometer. In

proteomics experiments, however, in-source fragmentation is rarely applied, and post-source fragmentation is the most commonly used. Post-source fragmentation involves the addition of energy to the ions after the ionisation/first mass analysis step. Several methods have been employed to transfer the energy required: transfer of electrons from a donor molecule (electron transfer dissociation [ETD] or electron capture dissociation [ECD]), exposure to electromagnetic radiation (photodissociation), and collision with a non-reactive gas or other molecule (collision-induced dissociation [CID] or higher-energy collision dissociation [HCD]). Each of the methods has its advantages and disadvantages, and the appropriateness of their use is dependent on the type of ions being fragmented (size, charge, modifications, etc.) and the instrumentation available. Moreover, the methods often provide complementary information, especially with regards to the analysis of post-translational modifications and can therefore be used in tandem or hybridised to maximise the information output (e.g. EThcD and decision-tree fragmentation methods). These methods are discussed with respects to phosphopeptide analysis in Section 1.6.3.

Collision-induced dissociation has been the most popular method for use in peptide sequencing in the last decade, and involves the excitation of selected ions by collision with a neutral gas (usually He, N, or Ar)[152,153]. As the selected ions collide with the gas, the energy of the collision is internalised, and the precursor ions begin to vibrate at a higher frequency and fragment (Fig 1.13A). The covalent bonds that break first are those that possess the lowest enthalpies, and at the collision energies often used in proteomics experiments (∼30-35 eV), these are the amide bonds that form the peptide backbone. Thus, in CID experiments, depending on where the positive charge remains (either on the carboxyl or amino group), two types of daughter ions (predominantly) form: *b*- and *y*-ions (Fig 1.13)[154]. HCD is very similar to ordinary CID; however, it employs a higher voltage to trap fragment ions within the C-trap and thus allow their mass analysis within an orbitrap. This results in an increase in the accuracy, mass range and resolution of the measurement of these ions. Moreover, this allows the measurement of low-mass ions which is not possible in an ordinary ion trap-CID experiment[155].

MS^2 spectra obtained from the fragmentation methods discussed above can be utilised for the sequencing of peptides without prior knowledge of their sequences – so-called *de novo* sequencing. This is performed (often manually) by calculating the difference in *m/z* between sequential peaks of the same ion type in the spectrum (usually *b*- or *y*-ion series), as by definition, the $\Delta m/z$ between two peaks in an ion series represent the loss of a single amino acid from the peptide sequence. Some fragmentation methods have biases towards creating more *b*- than *y*-ions and *vice versa*; however, as the two ion series run anti-parallel to each other along the peptide backbone (*b* running from N $\rightarrow$ C terminus, and *y* from C $\rightarrow$ N), spectra possessing both series give greater confidence in the determined sequence of the peptide[152–154,156]. A number of other ion types are also produced as peptide ions are fragmented (e.g *a*- and *x*-ions corresponding to fragmentation of the C-C bond prior to the amide bond); however, *b*- and *y*-ions are the most predominant in scans acquired using CID-based methods.

Although arguably the most accurate way of determining a peptide's sequence from MS^2 data, *de novo* sequencing is very time-consuming and thus not suitable for large-scale, global proteomics experiments, although efforts to automate this process have been initially successful (e.g. PEAKS software)[157]. As an alternative, MS^2 spectra can be processed and condensed computationally to yield raw peak lists, which can then be searched against a protein/peptide database of known sequences and predicted tryptic (or other protease) sequences calculated from *in silico* digests[158]. Amongst other similar methods, this can be achieved by first matching the parent ion's *m/z* from the MS^1 data to peptides within the database that could share that *m/z* – usually at a high accuracy (e.g. $\pm$10 ppm). As current instruments do not possess the mass accuracy re-

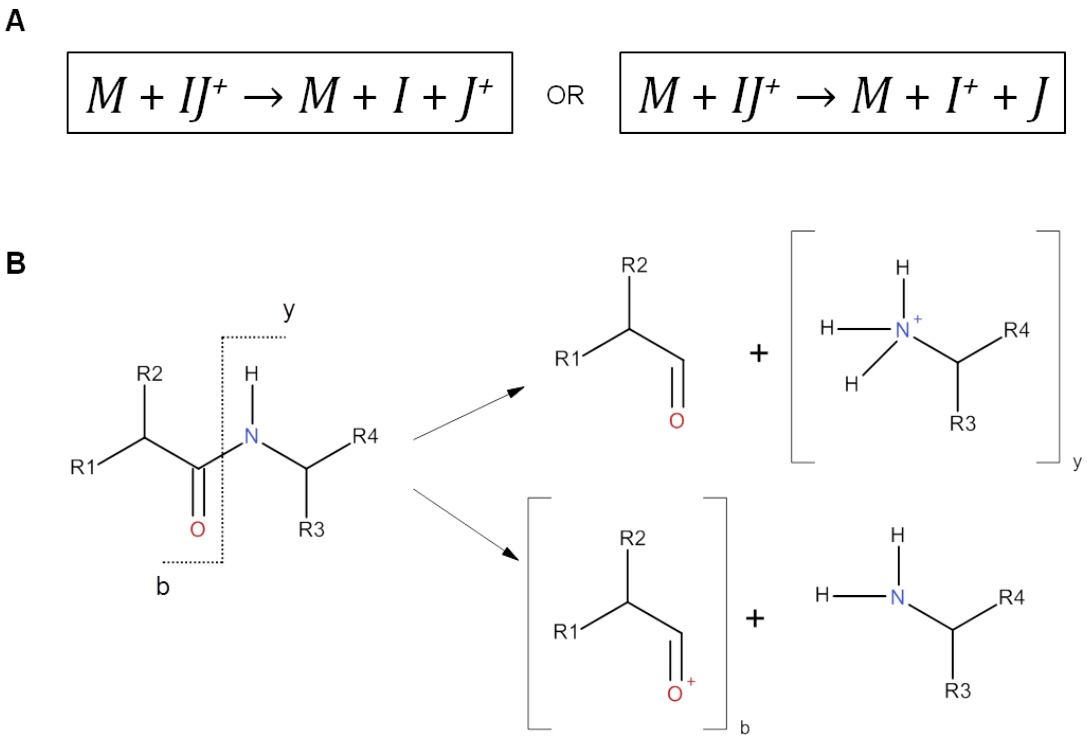


Figure 1.13: The principles of collision-induced dissociation. (A) A simplified equation demonstrating the process of collision-induced dissociation of peptide precursor ions. M = neutral gas molecule; IJ^+ = positively charged precursor ion; I = neutral fragment ion; J^+ = charged fragment ion. (B) Skeletal structures demonstrating the formation and differences between *b*- and *y*-ions. If the positive charge remains on the nitrogen it is noted as a *y*-ion, if it remains on the double-bonded oxygen it is noted as a *b*-ion. The two ion types produce ion series (multiple ions representing the incremental losses of a single amino acid) that are anti-parallel to one another.

quired to unambiguously assign peptide IDs from their parent ion *m/z*, peptides sharing the same *m/z* value are then scored according to how well the MS^2 spectrum peaks match those that would theoretically be produced from each of the potential peptide sequences[159]. Many different software packages are now available that have this capability, each with their unique way of scoring peptides; however, the most popular are arguably Mascot, Sequest, X!Tandem, and Andromeda. Mascot generates probabilistic scores based on the probability that the peptide match given by the algorithm occurred by chance. The final score is expressed as the $-10 \cdot log_{10}(P)$ where P is the probability score previously mentioned[159]. The main advantage of this database-searching method is the speed at which large MS datasets can be searched and peptide identifications made. However, due to the size of the search-space (the vast number of possible sequences and post-translational modifications) and the size of the datasets, the number of type I errors (false positives) can be substantial. Due to the flexibility of the scoring algorithms used, the user is able to set which *m/z* and significance scoring thresholds they are willing to accept; however, this can still sometimes result in a large number of false positives. To further reduce this, many software packages now include the ability to generate a false-discovery rate (FDR), with the aim of reducing the rate of type I errors and increasing the confidence in the qualitative data[160,161]. This process usually involves searching the data against a scrambled or reversed protein/peptide database to estimate the number of hits that are false positives. Other methods also take into account other parameters; the accuracy with which the precursor mass was identified, for example. The overall database of identified peptides can then be reduced to include a FDR below a certain percentage – this usually being set to between 1 and 5%.

Independent of the fragmentation and computational identification methods used, there are two main methods by which MS^2 scans are triggered: data-dependent acquisition (DDA) and

data-independent acquisition (DIA). DDA involves the process outlined in Fig 1.12, whereby precursor ions are scanned in MS^1, and then based on particular criteria, a distinct population of these ions is selected and fragmented successively in a semi-random manner. This cycle is then repeated for the length of the MS run. For example, the top 10 most intense parent ions from the MS^1 scan (at $t_R = x$) may be selected for fragmentation and potentially identified; however, the remaining precursor ions are ignored by definition (often the least abundant ones), leading to the issue of under-sampling. Several factors influence the rate of under-sampling in a DDA-MS run: instrument sensitivity, precursor ion intensity and isolation, and the instrument duty cycle (how quickly it can complete a full MS^1 and subsequent MS^2 scans)[162]. Although it remains a pertinent issue with DDA methods, the problem of under-sampling is constantly being reduced with advancements in instrument technology, which are consistently improving the duty cycles and sensitivity of current mass spectrometers.

DIA methods, developed to circumvent the impact of under-sampling on MS experiments, involve the fragmentation and mass measurement at the MS^2 level of all the parent ions present in the MS^1 scan. Thus, unlike DDA, the collected MS^2 spectra represent a composite of all the precursor ions that were fragmented within a particular mass window. Theoretically then, sequence information is obtained for all the parent ions present in a sample and thus a digital representation of all the proteins present in the sample is produced. Two methods that have been particularly successful in the pursuit of DIA are MS^E and SWATH. MS^E utilises a Q-TOF instrument to fragment all the precursor ions present in an initial MS^1 scan, while cycling between low and high collision energies[163]. Based on a similar principle, SWATH, involves the cycling of the instrument through multiple *m/z* windows over the entire chromatographic run time and fragmentation of everything within those windows at each particular t_R[164]. Therefore, similar to MS^E, every ion in the sample is subjected to fragmentation and thus the resulting digital file theoretically contains a record of every peptide in the sample. The peptides within the sample are then identified through reference to peptide spectral libraries produced via DDA methods – these ever increasing in size and public availability. These methods present multiple advantages, the most prominent of which being the depth with which the proteome can be analysed (as everything is fragmented). Furthermore, as the bias inherent in DDA methods is removed, there are much fewer missing data-points where peptides were not identified[165]. Lastly, these methods allow retrospective analysis of the data, allowing datasets to be interrogated in the context of new biological information as it becomes available. Alongside these benefits, there are distinct limitations to these methods. Firstly, the volume of data produced is orders of magnitude larger than that produced via DDA methods, and therefore data storage is limited by technological and monetary capabilities. Secondly, the data-mining and peptide identification algorithms rely on spectral libraries produced from DDA data, and therefore, although these are now sizeable and easily available, they do not yet represent the entire human proteome. Lastly, due to its unselective nature, the MS^2 spectra produced through these methods are often convoluted, especially with the wide isolation windows required to sample the entire proteome in a reasonable chromatographic time-frame. As a result, the confident identification of peptides from such spectra is a computationally demanding challenge. Despite these, numerous technological advances in recent years have continuously improved DIA methods and increased their availability to the proteomics community, particularly with regards to cloud computing and open-source softwares[166–168]

1.5.6 MS analysis – quantitative proteomics

Qualitative proteomic analysis is of profuse value to biological research; however, the ability of mass spectrometers to collect quantitative data is that of most worth to proteomics experiments and our understanding of biochemical processes. The methods and technologies that have been developed for quantitative analysis of proteins by mass spectrometry are grouped into classes: label-mediated and label-free, both discussed in more detail in the subsequent sections.

1.5.6.1 Labelling methods

Labelling approaches often involve the attachment of tags of known, differential masses to the proteins of interest. The protein samples to be compared are then mixed – at different stages of the workflow depending on the method being used – and analysed by LC-MS/MS. This process is summarised in Fig 1.14. The samples can then be distinguished on the basis of their differential masses; the differences being attributable to the known masses of the tags.

One of the first labelling methods to be developed, isotope-coded affinity tagging (ICAT), involves the post-harvesting labelling of proteins with either a "light" or "heavy" version of the tag. The tag consists of a reactive group (which reacts with a specific amino acid), an isotope-coded linker region, and a biotin affinity tag (Fig 1.14A). The protein extract of interest is incubated with a tag, digested with a protease, affinity purified to extract only the labelled proteins, and the resulting samples analysed by LC-MS/MS[169]. The difference in mass between the two "heavy" and "light" tagged samples is then resolved at the MS^1 level. Several other post-harvest labelling methods have also been developed: isobaric tags for relative and absolute quantification (iTRAQ)[170], tandem mass tags (TMT)[171], and dimethyl labelling[172]. Unlike ICAT, these methods involve labelling at the peptide level, post-digestion. Identical to ICAT, dimethyl labelling involves the quantification of peptide abundances at the MS^1 level, based on a known mass difference between heavy, medium and light peptides. iTRAQ and TMT, however, involve quantification at the MS^2 level. The latter relies on the production of characteristic reporter ions (of known mass) that are released from the peptides as they undergo fragmentation. The relative abundances of these reporter ions can then be used to compare the quantitative differences of the fragmented peptide between samples.

Despite their popularity and success, the aforementioned labelling methods have several limitations. Firstly, the efficiency of labelling can vary in a manner that is dependent upon the labelling conditions, protein/peptide abundance, and label manufacturer; all of which introduce significant variability in the quantitative measurements made[173]. Secondly, recent studies have suggested that the addition of tags to peptides increases the number of protons they accrue in the electrospray process, thus reducing their fragmentation efficiency and therefore the number of useful sequence ions produced[174]. Thirdly, peptides have to be identified to allow their quantification, thus significantly reducing the depth of analysis due to under-sampling. Lastly, and perhaps most pertinently, co-isolation of precursor ions selected for fragmentation has been shown to influence the quantitative accuracy and precision of MS^2-level based methods (as two ions of similar mass are fragmented, contaminating both the qualitative sequence information and the ratios of reporter ions)[175]. This has been improved with additional fragmentation steps (i.e. MS^3) to confirm the identity of the precursor ion; however, this adds time to already relatively lengthy analysis times[176]. Several distinct advantages of these latter methods exist, however, such as: their analytical simplicity, availability, and because they do not rely on MS^1 data, the instrument can run with a high top *N* spacing (the number of ions chosen to be fragmented between MS^1 scans) therefore reducing the impact of under-sampling.

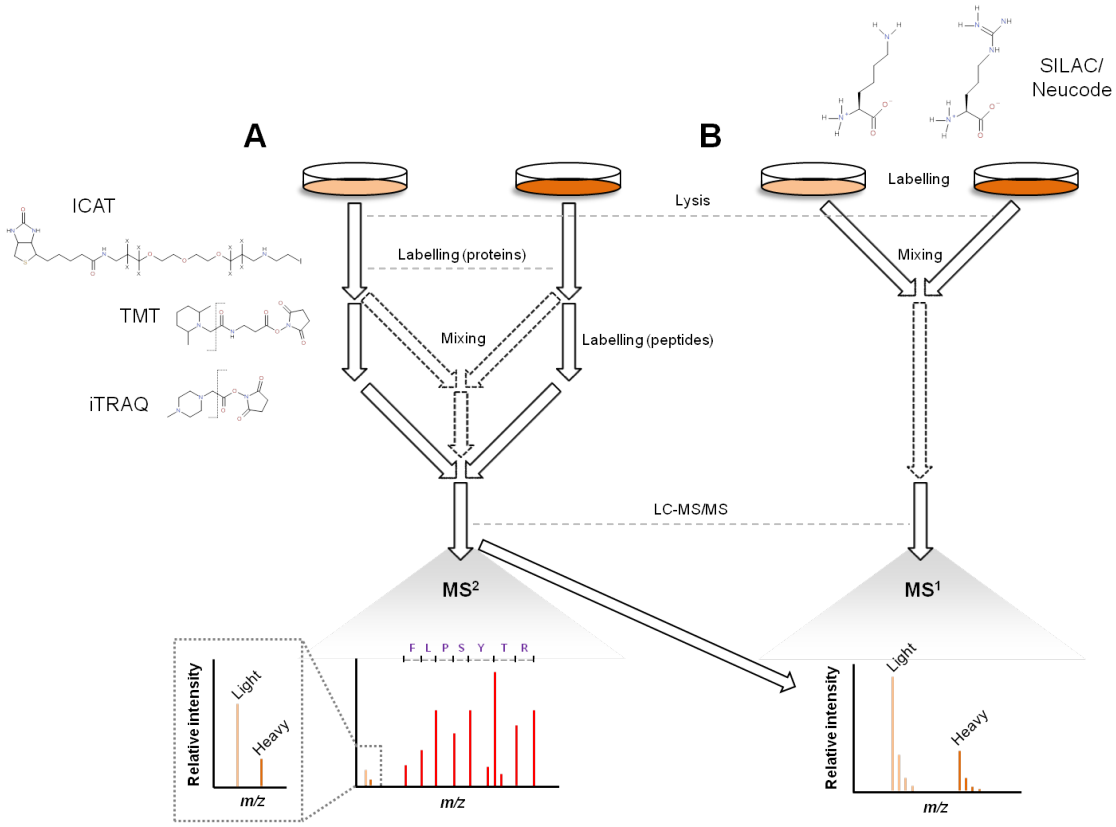


Figure 1.14: An overview of the labelling methods used for quantitative proteomics. (A) Samples are prepared independently and are labelled either at the protein (e.g. ICAT and dimethyl labelling) or peptide level (e.g. TMT and iTRAQ). The samples are then mixed, processed and run on the mass spectrometer. The relative differences in protein abundance between the samples being compared is then inferred at the MS^1 (e.g. ICAT) or MS^2 levels (e.g. iTRAQ and TMT). The latter relies on the dissociation of reporter ions from the isobaric tags following MS/MS fragmentation (shown with dashed lines with respects to TMT and iTRAQ labels). (B) Heavy and light isotopes of arginine and lysine are added to the growth media of the biological system of interest (SILAC and Neucode). As a result, proteins are labelled during synthesis, and upon mixing, digestion and LC-MS/MS analysis, the quantitative differences between the conditions can be inferred at the MS^1 level. ICAT = isotope-coded affinity tagging; TMT = tandem mass tagging; iTRAQ = isobaric tags for relative and absolute quantification; SILAC = stable isotope-labelled amino acids in cell culture; Neucode = neutron-encoded mass signatures.

Alongside the labelling of proteins after they have been harvested from the biological system of interest, an alternative approach is to label the proteins as they are synthesised *in vivo*. By far the most popular technique, termed stable isotope-labelled amino acids in cell culture (SILAC), involves the addition of heavy or light isotopes of lysine and arginine to the culture medium, allowing them to be incorporated into newly synthesised proteins as the cells grow and divide (Fig 1.14B)[177]. This removes the biases introduced from affinity purification and from labelling reaction kinetics. Moreover, as the samples are mixed as soon as the lysates are produced, any downstream variability in sample handling is removed. Despite this, there are numerous disadvantages with this technique. Firstly, the incorporation efficiency of each labelled amino acid can vary substantially. Secondly, some cells are unable to grow in heavy medium due to the dialysed foetal bovine serum required (as the dialysis of the FBS often removes important growth factors and other molecules). Thirdly, the technique is not applicable to primary tissue samples, as direct labelling is impracticable. Lastly, the number of different conditions that can be compared simultaneously is very small (typically 2-3), greatly reducing the number of biological questions that are tractable with this method. Neucode (neutron-encoded mass signatures) have been developed to overcome this latter limitation, by using super-resolution mass spectrometry (480,000 FWHM) to allow the resolution of labels that differ in mass over several different isotopic atoms[178]. Furthermore, several of the aforementioned limitations of SILAC have been improved through the use of so-called "Super"-SILAC, whereby SILAC-labelled cell-lines were utilised to generate tissue-specific proteomes that can be used as references for comparisons with other tissues/model systems[179]. Moreover, to counter the *in vivo*/whole-organism limitations of the method, a SILAC mouse was also developed by labelling mice over several generations to incorporate the amino acid labels[180].

Taken together, the different labelling methods discussed above, provide useful methods by which proteomes and sub-proteomes of different biological systems can be compared. The major limitations, however, are the lack of multiplexing afforded – although this is being improved with developments in the physicochemical properties of the labels being used – paired with the inability to apply the methods to clinical/*in vivo* samples, thus drastically limiting the applicability of the method to many biological questions.

1.5.6.2 Label-free methods

In contrast to labelling methods and to counter the limitations associated with them, a number of label-free methodologies have been developed for use in quantitative proteomics. As the name suggests, these methods rely not on the chemical labelling of proteins and peptides, but rather computational/analytical methods to make quantitative comparisons between sample proteomes (summarised in Fig 1.15A). The first of such methods, spectral counting, involves the use of the number of MS^2 scans a peptide triggers to infer its relative abundance in the sample; i.e. the more scans triggered, the more abundant the peptide. This method is particularly useful due its simplicity; however, the accuracy of the quantitative output is particularly susceptible to several confounding factors which have lead to the method often being described as semi-quantitative[181,182]. The most pertinent of these is the variability observed in peptide detection probabilities; this being dependent upon the physicochemical properties of the peptide in question. Moreover, it relies on the phenomenon of under-sampling as, by definition, it correlates the MS^2 selection bias with peptide abundance. Several methods to improve spectral counting data have been proposed and one such method, the Absolute Protein EXpression (APEX) algorithm, was designed to incorporate these confounding factors into the analysis by correcting the count data for variations in peptide

detection probability (which can be predicted from *in silico* digests)[183].

In contrast to spectral counting, a number of label-free methods utilise extracted ion chromatograms (XICs) and the respective height/area of the peaks generated to infer protein/peptide abundance. XICs are constructed by monitoring the abundance of a particular ion and often its isotopes – selected either computationally or physically in the mass spectrometer – over multiple scans, as a function of retention time (Fig 1.15B). Selective reaction monitoring (SRM) or multiple reaction monitoring (MRM) is a targeted label-free method where specific precursor ions of known mass are selected, fragmented and the abundance of their specific fragment ions monitored[184,185]. This method, usually performed in triple quadrupoles (QqQ), is very specific (given that the fragment ions/transitions are known) and very sensitive. Due to its specificity, the method affords short run-times and therefore relatively high throughput. As a result of this however, only a selection of analytes can be quantified in each run, although this has been improved through the use of scheduled SRMs. Scheduling allows the more efficient multiplexing of SRMs, as the instrument is set to only scan for a particular mass and/or fragment(s) at that analyte's known retention time. This drastically increases the overall number of analytes that can be measured in a single run. In a similar vein, parallel reaction monitoring (PRM) allows the monitoring of multiple transitions simultaneously, as the third quadrupole is substituted for a high-resolution mass analyser (e.g. an orbitrap)[186].

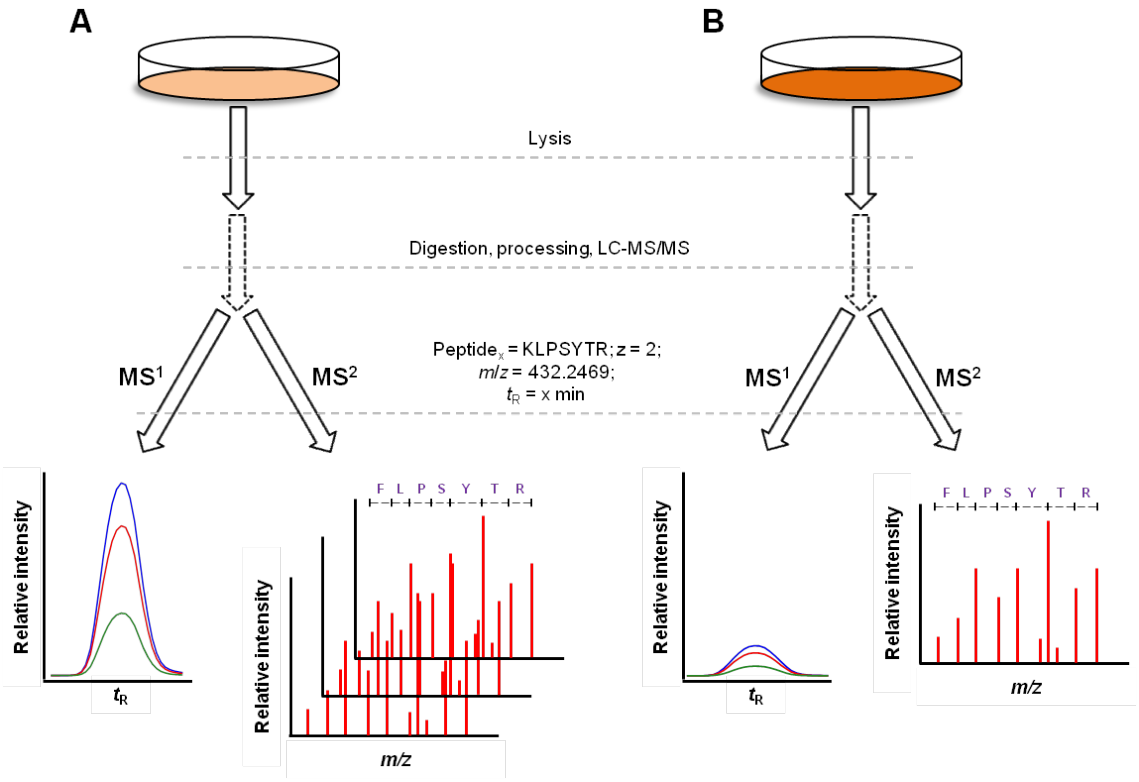


Figure 1.15: An overview of the label-free methods used for quantitative proteomics. In label-free methodologies, samples are lysed, processed and run on the mass spectrometer independently. After having identified the peptides by means of MS^2 data, the peptide can be quantified at the MS^2 level (e.g. spectral counting, where the number of times a peptide is fragmented is used to infer its relative abundance) (A) or at the MS^1 level (e.g. XIC targeted quantification, whereby the targeted peptide's abundance is inferred from its elution peak height or area (B). Blue, first isotope [all ^{12}C]; red, second isotope [1 x ^{13}C]; green, third isotope [2 x ^{13}C]). The latter method is advantageous as, by using the accurate *m/z* and t_R data, peptide XICs can be constructed and quantified even in samples where the peptide was not subjected to fragmentation and identified.

Another label-free method, PESCAL (PEak Statistic CALculator) – developed in the Cutillas lab and similar to the accurate mass tag (AMT) approach described by the Smith lab[187] – also constructs MS^1 XICs to quantify peptides/proteins, but in a more high-content manner. To

achieve this, a central database of peptides is constructed from the MS^2 data gathered from the experiment in question (or other experiments run on the same instrumentation). To account for chromatographic variations, the retention times of peptides common to each sample (or peptides from an internal standard protein) are calculated. These retention times are then used to form a linear model between each sample (Fig 1.16A). The quantification of peptides then requires several computational steps. First, the peptide to be quantified is located within the database and its *m/z* (± 5-7 ppm), charge state (*z*), t_R (± 1-2 min), and sample within which it was identified (e.g. $sample_x$) are stored. PESCAL then creates XICs at that specific *m/z* and t_R in the sample(s) in which the peptide is to be quantified (e.g. $sample_y$ – Fig 1.16B). In order to account for chromatographic variability, the t_R window at which it creates the XIC is sample-specific. These predicted t_R values are computed through interpolation of the linear models formed between the peptide standards in each sample (Fig 1.16A). In this manner, a peptide can be quantified in a sample in which it may not have been selected for MS^2 and identified (as a result of under-sampling). To increase the specificity of this peak-picking, the ratios of the second and third isotopes are correlated with the theoretical distributions for the peptide being quantified. Moreover, in order to improve the accuracy of the quantification, only peaks with a minimum number of data-points are accepted (to ensure accurate peak shapes). Following the picking of the peak of interest and assuming it adheres to the necessary criteria, the height or area of the peak is calculated and taken for further analysis[188].

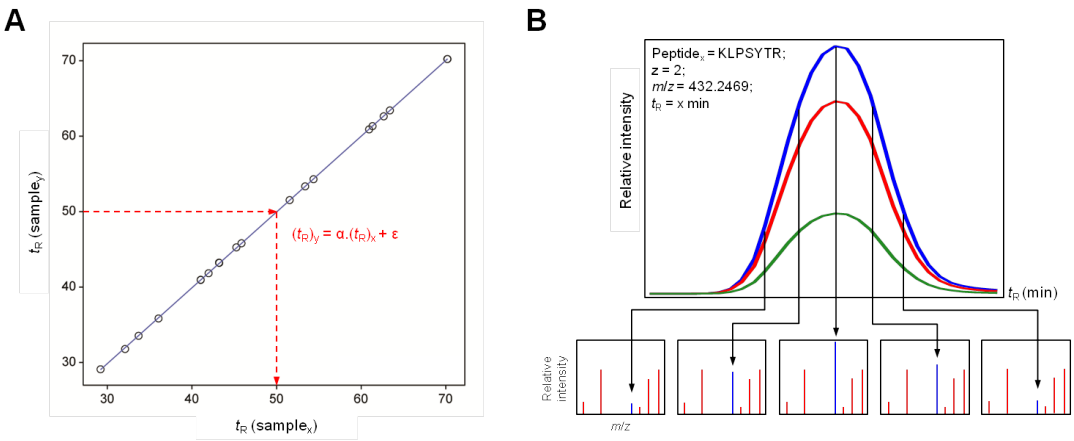


Figure 1.16: An overview of the PESCAL method of label-free quantification. (A) The retention times of known peptide standards (either endo- or exogenous) are aligned for each sample combination. A linear model is formed between the enolase standard retention times in the sample in which the peptide was identified ($sample_x$) and the sample within which the peptide is being quantified ($sample_y$). This model is then used to predict the retention time of a peptide identified in $sample_x$ in the sample being quantified (in this case $sample_y$) – red arrows. (B) XICs are constructed from the MS^1 data for the first (blue), second (red), and third (green) isotope of each peptide in the central database formed from all the samples in the experiment (or others), within user-defined *m/z* and t_R windows.

The main advantage of this approach is its high-content nature and its ability to compare a theoretically unlimited number of samples, without needing to identify the peptide in every sample. In this regard, it helps to overcome the issue of under-sampling in DDA methods, increases the scope of the biological questions that are tractable with the method, and is easily applied to clinical/*in vivo* samples. Alongside these however, are several disadvantages. Firstly, the specificity of the peak-picking. As the peptide may not have been identified in the sample in which it is being quantified, there is a certain margin of error in the assignment of an MS^1 peak to a particular peptide. The increased availability of very high mass-accuracy instruments (sub-ppm), more robust liquid chromatography, and the added layer of isotope distribution correlation has rendered this less of an issue than it once was however. Moreover, the increased availability of ion mobility cells provides the potential to add a fourth dimension of specificity (i.e. ion drift time). Secondly,

as with all label-free methods, each of the samples must be prepared independently, therefore increasing the length of the preparative work-flow and introducing potential sources of variation. This is compensated for by the ability to compare as many conditions as the experiment requires however, and in the same vein, many biological and technical replicates can be included in the experimental design to account for this. Despite this, many studies published from this and other groups have demonstrated the accuracy and reproducibility of the method[27,97,98,125,188–193], one study reporting an average CV of 34% and error of 6%[194], and work from the Cutillas lab demonstrating an average CV of 22%[141]. A number of other softwares from several labs, utilising similar principles, have recently been developed and made freely available (e.g. MaxLFQ and Skyline from the Mann and MacCoss labs respectively)[195,196].

1.5.6.3 Absolute quantification in proteomics

Regardless of the methodology employed, the large majority of proteomics experiments performed only require relative quantification of peptides and proteins in order to draw biologically relevant information from the data. In contrast to this, however, techniques exist for the absolute quantification of peptides and proteins. These are of distinct interest to several fields of study; particularly with regards to diagnostic and forensic applications. Several absolute quantification methods have been developed, involving both labelled and label-free strategies, and these have been applied to numerous biological systems.

Label-free methods for absolute quantification include: APEX (absolute protein expression profiling), TOP3, iBAQ (intensity-based absolute quantification), and the "proteomic ruler" method. APEX, as previously mentioned is based on spectral counting coupled to a probabilistic model of peptide observation frequencies. Contrary to this, the TOP3 method utilises the mean or sum of the MS^1 intensities of the three most intense peptide ions corresponding to a protein. Similar to the TOP3 method, iBAQ also utilises the intensities of the observed peptides; however, this method instead uses every observed peptide and normalises this to the total number of theoretically observable tryptic peptides for a particular protein. Each of these methods has been applied successfully in multiple studies, and their utility has been reviewed extensively[126,197–199]. A more recent method, termed the "proteome ruler", is based on the concept that the MS-signal of histone proteins is proportional to the amount of DNA and therefore the number of cells present in the sample. As a result, the absolute amount of a particular protein can be estimated from the relationship between its MS signal and the MS signal of the histone proteins[200].

Labelled strategies for absolute quantification are generally more complex, and include: absolute quantification (AQUA), protein standard absolute quantification (PSAQ), quantification concatemer (QconCAT), and protein epitope signature tags (PrEST), amongst others. Stable isotope-based methods have been the staple for the absolute quantification of a range of analytes for many years. In the context of proteomics, this involves the addition/chemical synthesis of a stable isotope-labelled version of the peptide/protein (e.g. an AQUA peptide or full-length PSAQ protein) of interest directly to the sample, and subsequent comparison of its abundance to that of the endogenous analyte[201,202]. As the absolute amount of synthetic peptide/protein that is added is known, the ratio of the synthetic and endogenous signals can be used to calculate the absolute amount of the analyte. Although highly accurate and conceptually simple when applied to a single protein, this method is less scalable than label-free methods, primarily due to the cost and validation of the synthetic peptide required. Despite this, some studies have used this method – or variations thereof – to construct a standard curve of protein abundances (e.g. using 19 proteins spanning 2 orders of magnitude) which can then be used to interpolate the absolute abundances

of the remaining proteome[203]. The QconCAT method circumvents a number of the issues present when using stable isotope-labelled standards, allowing more economical, multiplexed analysis. To achieve this, a synthetic protein is designed to contain concatenated proteotypic peptides (with predicited high ionisation efficiencies) corresponding to all the proteins of interest. This protein is then expressed in a vector system and labelled with stable isotopes. The protein can then be spiked into the test sample at a known amount and, when digested, yields heavy-labelled internal standard peptides that can then to be used for absolute quantification of the test proteins[204,205]. More recently, the PrEST approach, an extension of absolute SILAC methods, has also been applied to the multiplexed absolute quantification of proteins. This approach relies on the library of peptides selected and produced by the Human Protein Atlas Project, which represent approximately 100 amino acid-long sections of nearly 18,300 human genes[206]. These peptides are purified via their His tags, absolute amounts determined by SILAC, and are then spiked into test samples where they can be used to quantify the proteins they represent[207].

Absolute quantification of proteins possesses a clear role in proteomics, particularly with regards to clinical applications/diagnostics. Despite this, relative quantification methods, as previously discussed, are adequate for the large majority of biological experiments, particularly with respects to discovery experiments where large numbers of proteins are required to be compared quantitatively, and where time and cost are key considerations.

1.6 Sub-proteomics: the phosphoproteome

Numerous studies have now demonstrated that the expression of individual genes at the mRNA level correlates poorly with the expression of the proteins they encode[208–211]. On a similar train of logic, it has become increasingly apparent that the abundance of particular enzymes, in most cases, does not correlate with their actual activities[212]. In contrast, knowledge of their spatial distribution, post-translational regulation, and the abundance of the products of the reactions they govern are now appreciated as a more representative readout of their catalytic activity. As a result, the study of sub-proteomes – either in terms of sub-cellular distributions or specific populations of post-translationally modified (PTM) proteins – has become a popular area of research, particularly with the increasing accessibility of the MS technologies that are able to define these. Such studies often involve the specific enrichment of proteins possessing the PTM of interest (e.g. phosphorylation, acetylation, or glycosylation), isolation of a particular class of protein (e.g. kinases or histones), or purification of specific cellular organelles/compartments (e.g. mitochondria or Golgi apparatus).

Due to the long-known importance of protein phosphorylation in cellular signalling, its relatively simple and amenable chemistry, its relative abundance, and the development of ever more powerful MS and LC technologies, phosphoproteomics (the study of the entire complement of phosphorylated proteins in a biological system) has been perhaps the most studied sub-proteome in recent years.

1.6.1 The function of phosphorylation events

The human genome encodes for 516 distinct kinases (representing approximately 2% of total human genes), and approximately 160 of these have been implicated in disease[213]. Whilst other phosphorylated residues exist in nature (e.g. phospho-histidine and -arginine in prokaryotes), human kinases preferentially phosphorylate serine and threonine residues, tyrosine residues, or possess dual-specificity. The impact of phosphorylation on protein function is diverse in nature;

however, the main regulatory roles that have been demonstrated experimentally include: facilitation of protein-protein interactions, mediation of protein degradation and/or trafficking, regulation of protein thermodynamic stability, and control of enzyme activity. Often these functional consequences of phosphorylation events are intertwined. For example, dual-phosphorylation of Akt at Thr^{308} and Ser^{473} results in a substantial conformational change – presumably due to the thermodynamic considerations regarding the addition of two hydrophilic phosphates – which leads to the full activation of the enzyme[214]. Similarly, as previously discussed in Section 1.1.4, autophosphorylation of RTKs forms specific binding sites that recruit multiple signalling proteins (e.g. PI3K). Therefore, as a consequence of their chemistry, phosphorylation events play a key role in the propagation, amplification, and correct interpretation of signals from both intra- and extracellular sources and these are pivotal to the correct functioning of cellular and organismal systems.

On the contrary, however, some argue that not all phosphorylation sites (especially those that are not evolutionarily conserved) may have no function at all[215–217]. Much like the paradigm of "junk" DNA following the initial sequencing of the human genome however, it is more probable that this is a result of an absence of evidence as opposed to a true evidence of absence.

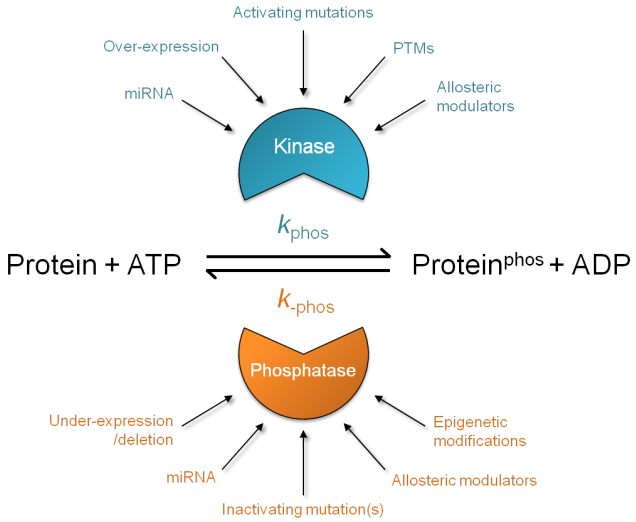


Figure 1.17: Phosphorylation sites are governed by kinase-phosphatase pairs. The kinase, utilising available ATP, catalyses the phosphorylation of the substrate protein (at a rate of k_{phos}). The phosphatase catalyses the removal of the phosphate group from the protein (at a rate of k_{-phos}).

The occupancy of individual phosphorylation sites is governed by the kinases and phosphatases that phosphorylate and dephosphorylate them, respectively (Fig 1.17). The rate at which the kinases and phosphatases catalyse their reactions, and therefore the overall equilibrium between protein and phosphorylated protein, is determined by numerous external processes, including: epigenetic and genetic regulation, protein-protein interactions, activating mutations on the kinase, and miRNA, amongst others. Furthermore, numerous kinases and phosphatases may act on the same substrate protein, complicating this determination of k_{phos} and k_{-phos}. In these cases it is thought that the specificity of kinase/phosphatase-substrate relationships (i.e. which kinase is phosphorylating the substrate), can be stimulus-, cell-, and context-dependent. For example, endoplasmic reticulum stress has been shown to alter the way in which Akt is regulated (i.e. the ratio between Thr^{308} and Ser^{473}) and its substrate specificity, in a severity-dependent manner[218]. Similarly, other data suggest that Akt substrate specificity is also altered as a result of its compartmentalisation to the endosome by Appl1 upon growth-factor stimulation[219]. As the interaction between the kinases and phosphatases that regulate protein phosphorylation are so complex, it may be posited that the most accurate readout of their combined functions is through

the measurement of the abundance/concentration of the phosphorylated substrate. Although phosphophorylated proteins can be detected by means of immunochemical methods, mass spectrometry has been instrumental in the unbiased and global analysis of phosphorylation events, as tens of thousands of phosphorylation sites can be fully quantatitatively monitored in a single experiment. There are, however, several key factors that must be considered in order to achieve this.

1.6.2 Enrichment of phosphorylated proteins/peptides

Despite the fact that the large majority of proteins possess phosphorylation sites, the substoichiometric nature of most leaves them outside of the dynamic range of mass spectrometers when measuring the total proteome. Instead, in order to reliably detect phosphorylation sites, they must be selectively enriched from the total population of proteins representing a particular biological system. A variety of methods have been developed for this purpose, including chromatographic and chemical derivation strategies; however, the most popular and arguably the most efficacious methods that have emerged are those involving either inorganic or immunological affinity purification, and these are discussed in further detail in the following sections. Another key consideration with regards to phosphorylation sites is their biological and chemical stability in solution. Phosphate groups possess a relatively long half-life in solutions within acidic or slightly basic solutions (pH 1-11); however, at high temperature and high [OH^-], (e.g. pH $>$ 12) they can be removed from the protein of interest via β-elimination[220]. Furthermore, phosphate groups are very susceptible to enzymatic removal due to the robustness of phosphatases in chaotropic/lysis conditions and the high rate at which they function (approximately 10-100 times faster than that of kinases). As a result, samples being processed for phosphoproteomics experiments must be kept at a low temperature ($<$ 4 ^{o}C), in relatively acidic/neutral conditions, and in the presence of several phosphatase inhibitors prior to enzymatic digestion.

1.6.2.1 Chemical derivatisation

One of the earliest methods utilised to identify phosphopeptides was by means of chemical derivatisation. Perhaps the most commonly used technique involves the removal and substitution of the phosphate group(s) via β-elimination and Michael addition, respectively. β-elimination is achieved through the addition of a high concentration of NaOH (to provide [OH^-] ions) and in some cases barium salts (e.g. $Ba(NO_3)_2$) or $Ba(OH)_2$ [221,222]. Michael addition is then performed through incubation of the sample with a dithiolated reducing agent (e.g. dithiothreitol [DTT] or ethanedithiol [EDT]) in order to modify the exposed double-bond and to allow selective isolation of these from the biological matrix of interest[221,222].

Despite its inherent specificity towards phosphorylated peptides, this enrichment strategy presents several key disadvantages. Firstly, due to the large number of reactions (and their intrinsic inefficiency), substantial sample-loss is observed and significant variability is introduced. Moreover, due to the inefficiency of the reaction steps, a number of side-reactions often produce unwanted peptide modifications, although this is reduced through the addition of EDTA[221]. Secondly, the biotin-avidin interaction exploited for the affinity purification is very strong, rendering it difficult to remove the phosphopeptides once they have bound. Thirdly, the common addition of a maleimide group for the selective isolation is unstable under common MS^2 conditions, thus increasing the complexity of the resulting spectra. Finally, many of the reactions possess a preferentially higher rate for phospho-serines over phospho-threonines, and thus are biased towards

the analysis of the former. Despite these issues, such derivation methods have been successfully applied to analyses of both phosphopeptides and phosphoproteins[221–223].

1.6.2.2 On-line liquid chromatographic methods

Alongside the susceptibility of phosphate groups to hydrolysis, further physico-chemical properties can be exploited for their enrichment. For example, liquid chromatographic separations can exploit the pK_A, charge and/or hydrophobicity of phosphopeptides, depending on the sorbent used. To this end, several stationary phases have been used to selectively enrich for phosphopeptides, including strong cation exchange (SCX)/strong anion exchange (SAX), high-pH reversed-phase (HpH), electrostatic repulsion-hydrophilic interaction chromatography (ERLIC), and hydrophilic interaction chromatography (HILIC)[224–226]. SCX and SAX technologies separate peptides (or other analytes) based upon their charge. In each case, the more charged the analyte, the greater the affinity of its association with the sorbent; SCX binding cations with greater affinity, and SAX binding anions. Therefore, as phosphopeptides carry ≥ 1 extra negative charge, such sorbents can be used to preferentially bind phosphopeptides (SAX) or preferentially bind non-phosphopeptides (SCX)[227]. In practice however, although broadly successful at enriching for populations of phosphopeptides, due to the wide-range of charges present on phosphopeptides (i.e. non-phosphopeptides having the same number of positive/negative charges as phosphopeptides), both methods suffer from poor specificity. Recent data suggest that HpH chromatography may be superior to SCX due to its better fractionation of singly phosphorylated peptides[226]. Similar in principle to both SAX and SCX chromatography, ERLIC and HILIC-based sorbents have also been used to enrich phosphopeptides. ERLIC and HILIC stationary phases bind peptides in a fashion that is diametrically opposed to that of reversed-phase sorbents – i.e. it binds them by means of hydrogen-bonding, thus binding with greater affinity to more hydrophilic/polar analytes. In addition, ERLIC involves electrostatic repulsion of highly positively charged groups. Therefore, as phosphopeptides tend towards greater polarity due to the extra negative charge(s) on the phosphate group, ERLIC/HILIC can be used to preferentially select for phosphopeptides. Much like SAX/SCX, however, the use of ERLIC/HILIC for phospho-enrichment in isolation suffers from poor specificity, as other polar peptides (e.g. those with acidic side-groups) are concomitantly enriched[228].

In spite of their previously mentioned lack of specificity, these methods are very effective when used in conjunction with other enrichment methods (e.g. TiO_2) in an orthogonal manner. These methods, when used off-line in conjunction with an additional enrichment step and traditional on-line LC methods, are often referred to as two-dimensional LC (2D-LC). This process involves the separation of the original sample material via an LC method which is orthogonal to that used for the on-line separation (e.g. SCX/SAX/ERLIC/HILIC and RP, or two stages of RP at different pH), and subsequent fractionation of the resulting eluents. Such strategies, when coupled to state-of-the-art MS technologies have been shown to drastically improve the coverage and depth of analysis of the [phospho-]proteome in many biological systems[226,229,230]. This is mainly due to the large reduction in dynamic range with respects to each fraction, thus rendering phosphopeptide enrichment more effective for low abundance species. On the contrary, however, these methods also substantially increase analysis time (as many fractions are created from a single sample), thus reducing the number of samples – and therefore replicates – that can be compared. Furthermore, they are not amenable to many forms of label-free analysis as accurately tracking peptides across multiple fractions presents a significant computational challenge. Finally, these methods also suffer from significant sample loss due to the multiple chromatographic steps, thus

increasing the amount of starting material that is required for each experiment[231].

1.6.2.3 Immunological affinity purification

Much like traditional affinity purification of specific proteins, antibody-based affinity purification can also be applied to the enrichment of phosphorylated proteins[232]. In this, antibodies are raised to recognise and bind to pSer/pThr or pTyr residues. Although an attractive route through which phosphopeptides can be enriched, these methods are heavily reliant on the quality of the antibody used and, although many pSer/pThr antibodies exist, these have been demonstrated to not be particularly effective due to their lack of specificity. On the contrary, anti-pTyr antibodies have enjoyed more success, perhaps due to the larger spatial footprint of the pTyr residues increasing antibody specificity and the lower abundance of pTyr residues in the total phosphoproteome (representing approximately 1.8% of all phosphorylated residues)[233]. Furthermore, due to the lack of specificity of other enrichment methods for the retrieval of pTyr residues, those based on antibodies are the most commonly used and many pTyr-focused studies have employed them successfully[234–236].

1.6.2.4 Inorganic affinity purification

Similar to the aforementioned LC methods for the enrichment of phosphopeptides, the multiple, pH-dependent negative charges that phosphate groups possess can be utilised in order to selectively enrich them from the total proteome. This can be achieved by exploiting the affinity of phosphate groups to positively charged inorganic species ligated to solid matrices. The three most popular schema that have emerged for this class of enrichment technology are: metal oxide affinity chromatography (MOAC), immobilised metal ion affinity chromatography (IMAC), or a combination of the two.

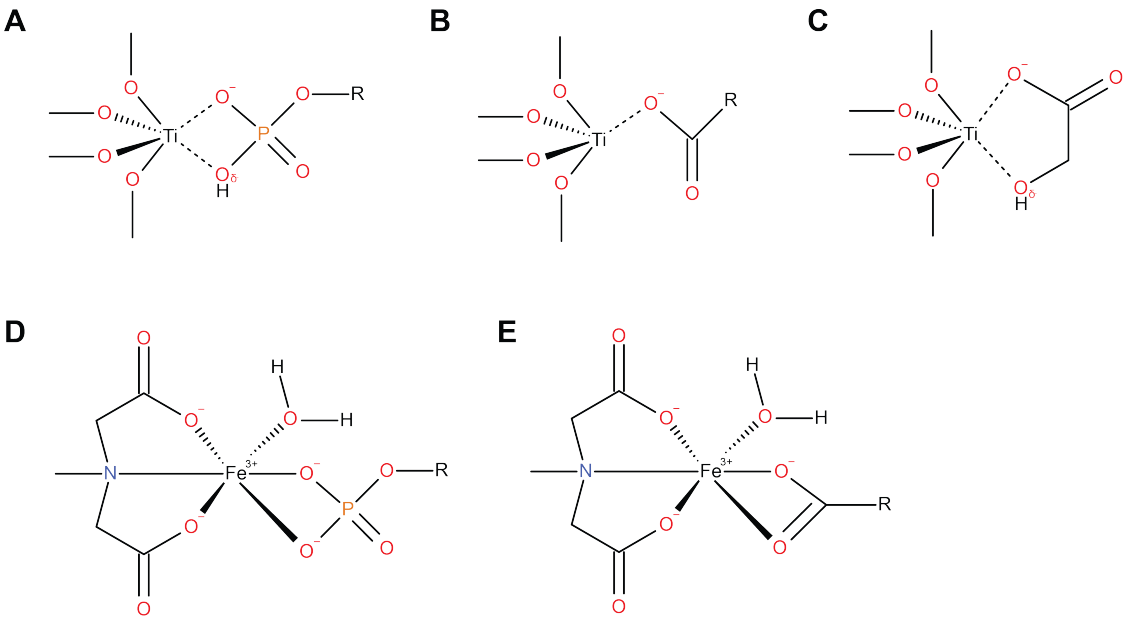


Figure 1.18: MOAC and IMAC methods for enrichment of phosphopeptides. (A) MOAC: phosphate groups bind to the TiO_2 sorbent via non-covalent interactions. (B) Acidic peptides also bind to TiO_2 through similar mechanisms. (C) α-hydroxy acids (AHAs), such as glycolic acid (shown) bind to TiO_2 with greater affinity than acidic peptides, but less affinity than phosphate groups, therefore increasing binding specificity of phosphopeptides by blocking empty binding sites. (D) IMAC: phosphate groups form complexes with immobilised metal ions. (E) Acidic peptides can also complex with the metal ions.

MOAC is based on the interaction between the partially and fully negatively charged oxygen atoms present within the phosphate groups and group 4 metal oxides (usually TiO_2 and ZrO_2), as shown in Fig 1.18A. Despite equal availability of ZrO_2, TiO_2 has been the medium of choice for the large majority of experiments involving phospho-enrichment and its utility has been frequently demonstrated and its efficacy drastically improved since its introduction in 2004[237]. Once the phosphate groups are bound to the sorbent, and following any washing steps to improve specificity of the enrichment, they can be eluted. Initial attempts to elute phosphate groups using phosphoric and other acids were superseded by the use of basic solutions (e.g. NH_4OH), as the latter have proved more effective[238]. The elution process involves the flooding of the sorbent with a high concentration of OH^- ions (which possess a higher affinity for the sorbent) in order to displace the phosphate groups.

Despite its relative success, TiO_2-based MOAC suffers from several limitations. Firstly, alongside its affinity for phosphate groups, other negatively charged moieties (e.g. the acidic amino acids Glu and Asp) also bind with relatively high affinity (Fig 1.18B). A number of chemical additives have been developed to circumvent this; for example, the addition of trifluoroacetic acid (TFA) has been demonstrated to drastically improve the selectivity for phosphate groups[238] as, by reducing the pH it causes Glu and Asp residues to become protonated. Similarly, the addition of α-hydroxy acids (e.g. dihydrobenzoic acid [DHB], lactic acid, and glycolic acid) has also been shown to reduce non-specific binding of acid groups, as these bind to the TiO_2 with an affinity between that of acidic and phosphate groups (Fig 1.18C)[238–240]. Secondly, TiO_2 has such a high affinity of phosphate groups, multiply phosphorylated peptides are often very difficult to elute once bound, resulting in a bias towards the recovery of singly phosphorylated species; however, several studies have shown the effectiveness of multiple elutions from the TiO_2 in order to recover as many phosphopeptides as possible[238,239]. Finally, non-phosphorylated peptides bind to TiO_2 through mechanisms other than the interaction between oxygen atoms and the titanium. For example, recent data have demonstrated that peptides bind in HILIC mode, and thus extra washes are necessary to remove these interfering peptides[241].

Similar to MOAC, IMAC employs the use of metals to bind to phosphate groups; however, unlike MOAC, it is based on the use of complexed metal ions as opposed to their oxides. Common transition metal ions that are used for this purpose include: Zn^{2+}, Co^{2+}, Al^{3+}, Ga^{2+}, Fe^{3+}, Cu^{2+}, Ni^{2+}, and more recently Ti^{4+}. When passed across the sorbent the negatively charged phosphate groups form complexes with the positively charged metal ions, and can then be eluted through modification of the pH or with a binding competitor. Fe^{3+} ions have been used most frequently for phosphopeptide enrichment, as they have been shown to possess high affinity for phosphate groups (Fig 1.18D)[242,243]; however, recent data from the Heck lab has demonstrated the superior efficacy of Ti^{4+}-based IMAC[244]. Similar to MOAC, IMAC suffers from a lack of specificity due to the concomitant binding of acidic peptides (Fig 1.18E). Moreover, the transition metal ions used possess affinity for sulphur and nitrogen-containing moieties (e.g. Cys residues), further reducing the phospho-specificity. The specificity of IMAC enrichment methods has however been improved through the optimisation of the loading and washing conditions. For example, it has been shown that an acidic pH, addition of formic acid, and presence of a high proportion of organic solvent (e.g. ACN) each act to improve phosphopeptide binding and reduce non-phospho interactions with the sorbent[245,246].

Alongside their use in isolation, both MOAC and IMAC methods have been employed in combination with each other and with other chromatographic technologies in an effort to produce a comprehensive phospho-enrichment method. Examples of such techniques include SIMAC (sequential elution from IMAC) and TiSH (TiO_2, SIMAC and HILIC), both developed by the Larsen

lab[247,241]. SIMAC involves the sequential elution of phosphopeptides from the IMAC sorbent with solvents of different pH and subsequent TiO_2 enrichment of these fractions in order to gain greater coverage of the phosphoproteome and to resolve mono- and multiply-phosphorylated peptides. Similarly, TiSH involves the initial enrichment of phosphopeptides using TiO_2, followed by SIMAC, and subsequent HILIC fractionation of the mono-phosphorylated peptides produced from the SIMAC separation. Both of these methods have been demonstrated to be highly sensitive and produce excellent coverage of the phosphoproteome; however, the resources and time required to carry out the sample processing steps reduce the method's applicability to large-scale studies.

MOAC and IMAC technologies have revolutionised the study of the phosphoproteome by mass spectrometry; however, the technique of choice with regards to single-shot enrichment of individual samples has been demonstrated to be TiO_2-based MOAC, due to its simplicity, speed, specificity and sensitivity. Consequently, the Cutillas lab has developed a TiO_2-based enrichment method for use in single-shot, label-free, quantitative phosphoproteomics experiments[141].

1.6.3 MS^2 ion activation methods for phosphopeptide ions

Concurrent with the technical challenge of enriching phosphopeptides/proteins from complex biological matrices, phosphopeptides also present a significant challenge with regards to their MS^2-based fragmentation and therefore identification. This manifests itself in the physicochemical properties of the phosphopeptides themselves, as they possess a lower positive charge than their non-phosphorylated counterparts (due to the negative charge(s) on the phosphate) and thus fragment less efficiently when subjected to CID. Furthermore, due to the relatively labile nature of the phosphoester bond, when subjected to CID, pSer/pThr phosphopeptides often undergo neutral loss, whereby uncharged phosphoric acid (H_3PO_4) is lost from the peptide and dehydroalanine or dehydroaminobutyric acid are produced. The resulting MS^2 spectra contain a very intense peak corresponding to the peptide minus the neutral loss of H_3PO_4 and comparatively sparse sequence information, rendering phosphopeptide identification challenging (Fig 1.19). Alongside the sometimes poor fragmentation of phosphopeptide ions in the gas phase, it is often difficult to accurately localise the phosphorylation site in the peptide sequence, particularly when multiple phosphorylation sites are present. As a result of these limitations, numerous ion fragmentation methods and computational approaches have been developed to improve the sequencing and accurate site localisation of phosphopeptides by MS.

In order to combat the prevalence of neutral loss in phosphopeptide MS^2 spectra, an extension of CID fragmentation was developed, termed CID-MSA (multi-stage activation). In this pseudo-MS^3 method, the phosphopeptide undergoes normal CID fragmentation; however, the theoretical *m/z* values corresponding to a potential neutral loss are automatically fragmented and the resulting spectra combined (Fig 1.19). Another fragmentation method, HCD, has also been employed for the sequencing of phosphopeptide ions in the gas phase, and due to the ions being measured at higher mass accuracy and resolution in an orbitrap mass analyser, the data provide more accurate sequence and site localisation information than normal CID. On the contrary, however, due to the slower speed of HCD fragmentation in most instruments, CID still provides richer datasets with regards to LTQ-Orbitrap instruments[248]. Alongside the traditional CID-based fragmentation methods, ETD has also been used for phosphopeptide identification. ETD works on the principle of the transfer of electrons from a donor molecule (often fluoranthene) to the substrate peptide in the gas phase, therefore resulting in its dissociation. ETD presents with the advantage that it provides comprehensive sequence information alongside the preservation of post-translational modifications and has been used successfully in multiple proteomics studies. It is, however, most

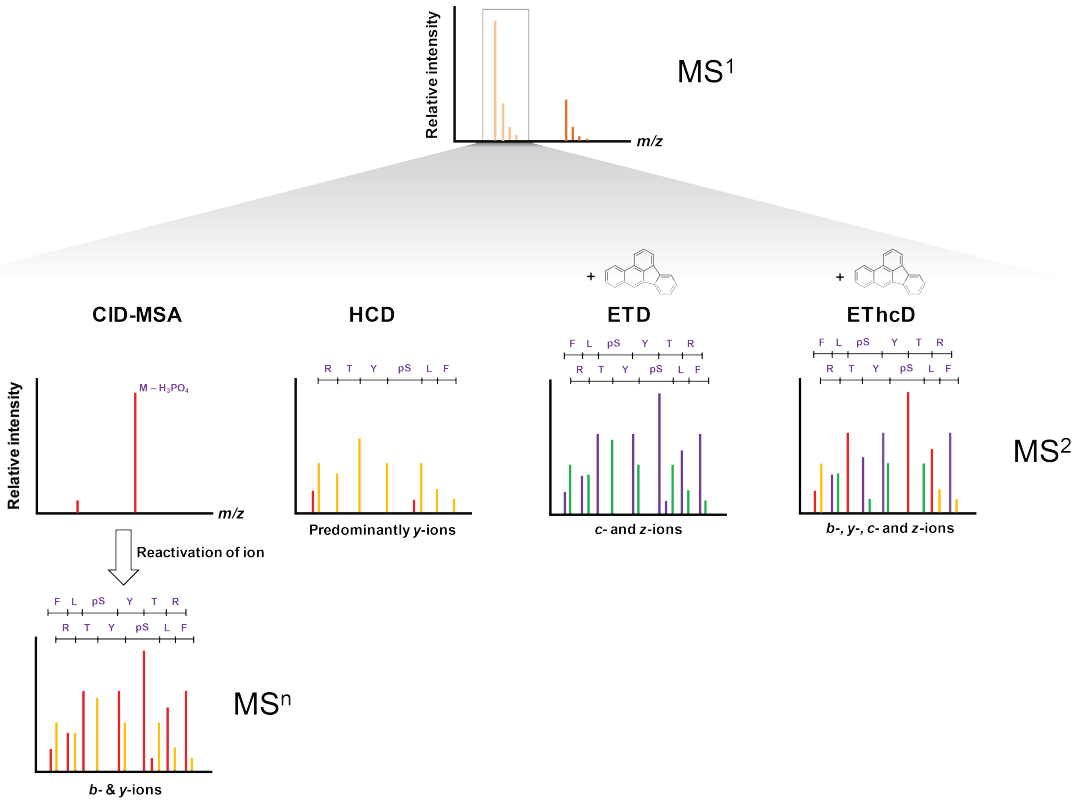


Figure 1.19: MS^2 fragmentation methods for phosphopeptide ions. Following data-dependent selection of a phosphopeptide precursor ion, a number of fragmentation methods can be used to confirm its sequence. CID-MSA (left) allows the activation of a an unfragmented phosphopeptide ion in a pseudo-MS^3 reaction, following neutral loss of H_3PO_4. These reactions primarily produce *b*- and *y*-ions. HCD (middle-left) produces data of higher mass accuracy and higher resolution compared to CID-MSA, as the precursor ion is fragmented in a separate octopole and measured in the orbitrap mass analyser, predominantly producing *y*-ions. ETD (middle-right) involves the reaction of the precursor ion with an electron donor (fluoranthene pictured), allowing fragmentation through electron transfer. ETD produces both *c*- and *z*-ions. EThcD (right), a hybrid between ETD and HCD involves the initial reaction of the precursor ion with an electron donor, and subsequent HCD fragmentation of the remaining ions. In this scenario, *b*-, *y*-, *c*-, and *z*-ions are producing, providing optimal sequence coverage.

effective when used with highly charged peptides ($\geq$ 3+); a feature largely incompatible with tryptic peptides, as these are commonly $\leq$ 3+. To circumvent this, it is possible to "super-charge" peptides within a sample through the addition of *m*-nitrobenzyl alcohol (*m*-NBA) to the mobile-phases of the LC system[128], thus making ETD more suitable for tryptic digests.

Analogous to the aforementioned phospho-enrichment methodologies, a combination of ion activation methods produces the best fragmentation of phosphopeptide ions. To this end, several methods have been developed which combine fragmentation technologies, either physically or computationally, to produce the most efficient sequencing of peptides. Firstly, EThcD, a physical hybrid of ETD and HCD technologies, allows the initial reaction of the precursor ion with fluoranthene to produce *c*- and *z*-ions, and subsequent HCD fragmentation of the remaining ions to produce further *b*- and *y*-ions[249]. Therefore, any intact precursor ion that persists following inefficient ETD fragmentation (perhaps due to the peptide possessing a low charge) will still produce fragments from the HCD stage and *vice versa*. Thus, the method compounds the advantages of both fragmentation methods and produces data-rich MS^2 spectra that allow accurate peptide sequencing, high peptide coverage, and high confidence PTM localisation[249,250]. Based upon a similar notion, other methods have sought to allow on-the-fly decision tree-driven fragmentation methods, whereby the instrument, based on the empirically determined parameters of the precursor ion (*m/z*, *z*, etc.), decides which fragmentation method to use on an ion-by-ion basis. For example, large ions possessing a high charge will be subjected to ETD, whereas smaller ions of lower charge will be subjected to HCD/CID[251]. Therefore, as each ion is subjected to the theoretically optimal fragmentation method, the best possible sequence information can be obtained in order to allow accurate peptide identification.

1.6.4 Bioinformatic considerations for the sequencing of phosphopeptide ions

In parallel with the improvements in ion activation technologies, there has been continued development of the search algorithms required to analyse the resulting datasets. The sequencing of phosphopeptide ions is a distinct computational challenge, primarily due to the prevalence of neutral loss events and the combinatorial complexity of the number of possible isomeric forms of any particular phosphopeptide (increasing exponentially with an increasing number of phosphorylation sites). These issues are compounded by the distinct peptide fragmentation patterns and probabilities between different ion activation methods. Thus, phosphopeptide search engines are required to produce an estimate of the error inherent in sequencing the amide backbone of the peptide itself (often referred to in the form of an FDR), whilst also provide an estimate of the error inherent in phosphosite localisation (or false localisation rate [FLR]). Several algorithms have been developed for these purposes, both proprietary and academic, including: Mascot Delta Score, PhosphoRS, and LuciPHOr. These algorithms, whilst differing in the exact method they use, each produce an objective assessment of the accuracy of the phosphosite localisation, and each has demonstrated its utility for phosphoproteomics experiments[252–254].

Chapter 2

Materials and Methods

2.1 Antibodies

Primary Target	Manufacturer	Product number	Specification	Concentration (v:v)
GADPH	Abcam	ab9485	Rabbit pAb	1:2000
Phospho-Akt (Ser^{473})	Cell Signaling	9271	Rabbit pAb	1:500
Phospho-ERK1/2 (Thr^{202}; Tyr^{204})	Cell Signaling	4370	Rabbit mAb	1:1000
EGFR	Cell Signaling	2232	Rabbit pAb	1:500
Phospho-MARCKS ($Ser^{152/156}$)	Cell Signaling	2741	Rabbit pAb	1:1000
Phospho-RPS6 ($Ser^{235/236}$)	Cell Signaling	2211	Rabbit pAb	1:1000
Phospho-S6K (Thr^{389})	Cell Signaling	9202	Rabbit pAb	1:1000
α-tubulin	Cell Signaling	2144	Rabbit pAb	1:1000
Vinculin	Sigma Aldrich	V9131	Mouse mAb	1:2000

Table 2.1: A summary of primary antibodies used for western blotting. pAb, polyclonal antibody; mAb, monoclonal antibody.

2.2 Miscellaneous reagents and materials

Reagent/material	Manufacturer	Product number	Supplementary information
Acclaim® PepMap RLSC C18 Analytical Column (75 μm x 25 cm, 2 μm, 100Å)	Thermo Scientific	164536	Analytical column used for Velos LC separations
Acclaim® PepMap 100 C18 Trap Column (100 μm x 2 cm, 5 μm, 100Å)	Thermo Scientific	164564	Trap column used for Velos LC separations
BEH C18 nanoACQUITY column (75 μm x 15 cm, 1.7 μm, 130Å)	Waters Corp.	186003543	Analytical column used for XL LC separations
Bradford assay reagent	Sigma-Aldrich	B6916	Protein concentration estimation
Bicinchoninic acid (BCA/Smith) assay reagent	Thermo Scientific Pierce	23227	Protein concentration estimation
CellTiter® 96 AQueous non-radioactive cell proliferation assay	Promega	G5421	Cell viability assay platform
C18 micro-spin columns	Protea Biosciences	SP-150-96	Phosphopeptide enrichment platform
C18 micro-spin columns	Glygen	TT2C18.96	Phosphopeptide enrichment platform
C18/graphitic carbon TopTip micro-spin columns	Glygen	TT2MC18.96	Phosphopeptide enrichment platform
Dulbecco's Modified Eagle's Medium (DMEM)	PAA/Sigma Aldrich	E15-843/D6429	MCF7 culture medium
Dulbecco's Phosphate-Buffered Saline (PBS)	PAA/Sigma Aldrich	H12-002/D8537	Used for all cell culture and lysate preparation
ECL HRP-linked anti-rabbit antibody	GE Life Sciences	NA934	Western blotting reagent
ECL HRP-linked anti-mouse antibody	Dako	P0260	Western blotting reagent
EDTA-trypsin (0.05%/0.02%)	PAA/Sigma Aldrich	L11-004/T4299	MCF7 culture reagent
Empty TopTip micro-spin columns (PE-filtered)	Glygen	TF2EMT	Phosphopeptide enrichment platform
Eppendorf LoBind micro-centrifuge tubes	Sigma-Aldrich	Z666505	Used for all sample preparation
Eppendorf LoRetention pipette tips	Fisher Scientific Ltd	PMP-907-040W	Used for all [phospho-]peptide sample handling were available
Foetal Bovine Serum (FBS)	PAA/Gibco	A15-104/10500-064	MCF7 culture medium supplement
Immobilon®-P PVDF membrane	Millipore	IPVH00010	Western blotting membrane
mini-PROTEAN® 12% PA-gels	Biorad	456-1044	PAGE reagent
mini-PROTEAN® 4-15% PA-gels	Biorad	456-1084	PAGE reagent
OASIS® HLB 1 cc (30/10 mg) cartridges	Waters Corp.	WAT094225	Reversed-phase solid-phase extraction cartridges
Penicillin/streptomycin	PAA/Sigma Aldrich	P11-010/P4333	MCF7 culture medium supplement
Phosphatase inhibitor cocktail I	Sigma Aldrich	P8340	Western blot lysis buffer supplement
PMSF	Sigma Aldrich	93482	Western blot lysis buffer supplement
PVDF membrane	Millipore	IPVH00010	Western blot component
Recombinant EGF (human)	Peprotech	AF-100-15	Cell culture medium additive
Recombinant IGF-1 (human)	Peprotech	AF-100-11	Cell culture medium additive
SuperSignal West Pico ECL substrate	Thermo Scientific Pierce	34080	Western blotting reagent
Symmetry C18 nanoACQUITY trap column (180 μm x 2 cm, 5 μm, 100Å)	Waters Corp.	186006527	Trap column used for XL LC separations
Titansphere TiO_2 beads (10 μm)	GL Sciences Inc.	5020-75010	TiO_2 chromatographic media
TPCK-trypsin (immobilised)	Thermo Scientific	20230	In-solution protein digestion reagent

Table 2.2: A summary of all miscellaneous materials and reagents used.

2.3 Kinase/phosphatase inhibitors

Compound	Manufacturer	Product number	Primary target	Concentration (μM)
Akt inhibitor VIII	Chemdea	CD0223	Akt/PKB	1
Bisindolylmaleimide-1	Calbiochem	203290	PKC	1
Cantharidic acid	Calbiochem	210150	PP2A	10
DG2	Calbiochem	559274	p70S6K	5
ERK inhibitor	Calbiochem	328006	ERK	40
ERK inhibitor II	Calbiochem	328007	ERK	30
GDC-0941	Chemdea	CD0245	PI3K (pan)	1
Gö-6976	Calbiochem	365250	PKC	1
GSK-1120212	Selleckchem	S2673	MEK	0.5
H-1152	Calbiochem	555550	ROCK	5
KN-93	Calbiochem	422708	CAMK2	5
KN-62	Calbiochem	422706	CAMK2	10
KU-0063794	Chemdea	CD0274	mTORC1/2	1
MK-2206	Selleckchem	S1078	Akt/PKB	1
Okadaic acid	Calbiochem	495604	PP2A	1
PD-153035	Calbiochem	234490	EGFR	1
PD-168393	Calbiochem	513033	EGFR	10
PF-4708671	Calbiochem	559273	p70S6K	20
PI-103	Chemdea	CD0113	PI3K (pan)	1
Torin-1	AxonMedchem	1833	mTORC1/2	1
U0126	Calbiochem	625005	MEK	10
Y-27632	Chemdea	CD0141	ROCK	10

Table 2.3: A summary of all kinase and phosphatase inhibitors used.

2.4 Cell culture – MCF7

All cell handling procedures, including lysis, were performed under an appropriate laminar flow hood using sterile materials. Parental MCF7 cells were routinely cultured using DMEM (10% FBS; 1% penicillin/streptomycin, unless otherwise stated), at 37 ^{o}C, in a humidified atmosphere containing 5% CO_2. The identity of the MCF7 cell-line was confirmed using short tandem repeat genotyping (AmpFlSTR® Identifiler® panel).

2.4.1 Cell seeding for experimentation

MCF7 cells were washed twice with PBS and subsequently detached through incubation at 37 ^{o}C (5 min) with 0.05% EDTA-trypsin. Inactivation of the trypsin was achieved through addition of 1:1 DMEM. The suspension was then centrifuged at 1,300 rpm for 5 min (18 ^{o}C) and the cells resuspended in fresh DMEM and counted using a haemocytometer or Beckman Coulter ViCell XR instrument. Cells were seeded 3 days before each experiment in the appropriate size tissue-culture treated plates, at approximately 7–9 x 10^3 cells·cm^{-2}. Each plate was washed once with PBS and fresh DMEM added approximately 24 hours before each experiment. For those experiments involving serum starvation, cells were washed twice with PBS and an appropriate volume of fresh serum-free DMEM (0% FBS; 1% penicillin/streptomycin) added 24 hours prior to the relevant experiment.

2.4.2 Cryopreservation

Cells were washed and detached from the relevant flask (as in Section 2.4.1), collected by centrifugation at 1,300 rpm for 5 min (18 ^{o}C), and resuspended in the appropriate volume of 10% DMSO; 90% FBS. The cells were then transferred to 1.8 mL cryovials, placed in isopropanol

tanks and transferred to -80 ^{o}C. After 24 hours, the cryovials were transferred to liquid nitrogen storage tanks (approximately -178 ^{o}C) for long-term preservation. When required, cells were removed from liquid nitrogen, thawed at room temperature and immediately transferred to fresh DMEM. The suspension was then centrifuged and resuspended in fresh DMEM.

2.4.3 Treatment with kinase/phosphatase inhibitors

Each compound was diluted to 1000 times the desired concentration for treatment (as Table 2.3), using dimethyl sulphoxide (DMSO). Each compound was then added to the relevant cell population at a 1:1000 dilution in the media. Cells were incubated with the appropriate concentration of the relevant compound for the indicated time prior to lysis.

2.4.4 Creation and maintenance of resistant cell-lines

Parental MCF7 cells were split into six identical populations of cells. Three cultures began GDC-0941 treatment, and three began KU-0063794 treatment at 100 nM. At each passaging of the individual cultures (when they reached approximately 80-90% confluency), the concentration of relevant drug was increased by 100 nM. This drug concentration was increased in this manner until a maximum of 1 μM of the relevant drug in the media was reached. Once at this point, the resistant cell-lines were maintained in 1 μM of the relevant compound at all times unless otherwise stated. The cell-lines resistant to each drug were maintained at the same passage number throughout.

2.4.5 MTS proliferation assays

The relevant cell-lines were seeded in 96-well plates (clear, flat-bottomed) at approximately 3,000 or 5,000 cells·well^{-1} (as indicated) in triplicate or quadruplicate. After 24 hours, the cells were then treated with the indicated concentrations of the specified drugs or vehicle (DMSO). Following a 48 hour treatment with the compounds, cell viability was determined with the MTS assay reagent (Promega, WI, USA), as per the manufacturer's instructions. The absorbance (490 nm) of each well was determined using a spectrophotometer – this measurement being acquired in duplicate. The ratio of absorbance of each concentration data-point versus the absorbance of vehicle treated cells was then calculated. Third-order polynomials were fitted to the data using the *ggplot2* package within the R statistical computing environment. IC_{50} values were then estimated by solving the respective fitted equations.

2.4.6 Crystal violet assays

The relevant cell-lines were seeded in 96-well plates at approximately 5,000 cells·well^{-1}, in triplicate. After 24 hours, cells were treated with the indicated concentrations of drugs or vehicle (DMSO). Following a 48 hour treatment with the relevant compounds, cells were fixed with 100 μL·well^{-1} PFA (4% w/v, PBS) for 30 min on ice. Following removal of the PFA, cells were stained with 100 μL·well^{-1} crystal violet (0.5% w/v, 20% MeOH; 80% ddH_2O) for 10 min at room temperature. Plates were then washed with water until no soluble dye remained. The crystal violet stain was then re-solubilised with 100 μL·well^{-1} Sorenson's buffer (0.1M $Na_3C_6H_5O_7$, 50% EtOH; 50% ddH_2O) and mixed on a plate shaker at room temperature for 30 min at 300 rpm. The absorbance (540 nm) of each well was determined using a spectrophotometer, in duplicate.

2.5 Cell lysis

Experimental use	Lysis buffer components
LC-MS phosphoproteomics/proteomics	**8M Urea in 20 mM HEPES (pH 8.0)** (supplemented with 1 mM Na_3VO_4, 1 mM NaF, 1 mM β-glycerol phosphate, and 2.5 mM $Na_2H_2P_2O_7$)
Western blotting	**50 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, and 1% Triton X-100 (pH 7.4)** (supplemented with 1 mM Na_3VO_4, 1 mM NaF, 1 mM β-glycerol phosphate, 2.5 mM $Na_2H_2P_2O_7$, 1 μM okadaic acid, 1X Sigma protease inhibitor cocktail and 1 mM phenylmethylsulphonyl fluoride [PMSF])

Table 2.4: A summary of lysis buffers used and their constituents.

Cells were washed with ice-cold PBS (supplemented with 1 mM Na_3VO_4 and 1 mM NaF). An appropriate volume of lysis buffer was added to each plate, the cells scraped, and the resulting lysate collected. Each lysate was then briefly sonicated, centrifuged at 20,000 x *g* for 10 minutes (5^{o}C) and the resulting supernatant kept for further analysis. The approximate protein concentration of each lysate was then determined using a standard Bradford or bicinchoninic acid assay (BCA) (utilising BSA to construct a standard curve – correlating absorbance at 595 (Bradford)/562 (BCA) nm with protein concentration).

2.6 SDS-PAGE

Lysates prepared for Western blotting analysis were normalised to an appropriate concentration using lysis buffer and an equal volume 4X Laemmli buffer (240 mM Tris-HCl *p*H 6.8; 40% glycerol, 0.02% bromophenol blue, 0.1 M DTT, 12.5% SDS). Samples were then heated at 95^{o}C for 5 min. Samples were loaded into 12% or 4-15% precast SDS-PA gels, the gels assembled into Biorad gel tanks, submerged in running buffer (25 mM Tris; 192 mM glycine; 0.1% SDS), and electrophoresis performed at 70 V for approximately 15 min and 100 V for 1 hour.

2.6.1 Western blotting

Following SDS-PAGE, proteins were transferred to polyvinylidene fluoride (PVDF) membranes (as Table 2.2) using the BioRad Trans-Blot® Turbo system as per the manufacturer's instructions. Transfers were performed at 25 V/1.0 A for 30 min. Once transferred, the membranes were blocked through incubation with blocking buffer (5% w/v skimmed milk powder in Tris-buffered saline with Tween (TBS-T: 0.15 M NaCl; 0.05 M Tris-HCl; pH 7.6; 0.1% Tween-20) for $\geq$ 1 hour. The membranes were then incubated at 4^{o}C with the relevant primary antibody (diluted in 5% w/v BSA; TBS-T – concentrations as Table 2.1) overnight. Prior to incubation with secondary antibody, each membrane was washed with TBS-T for 5 min (in triplicate). Membranes were incubated with secondary antibody (anti-rabbit/anti-mouse Ig conjugate diluted in 2.5% w/v skimmed milk powder in TBS-T) at room temperature for 1 hour. Following this incubation, membranes were washed with TBS-T for 5 min (in triplicate), excess solution removed, and incubated with 1X SuperSignal West Pico ECL solution (2 mL·membrane^{-1}) for 1 min. Antibody affinity was then visualised using X-ray film.

2.7 LC-MS sample preparation

2.7.1 Tryptic digestion

Prior to digestion, all lysates (normalised to 500 or 250 μg using lysis buffer) were reduced (10 mM DTT for 30 min at room temperature in the dark) and alkylated (16.6 mM iodoacetamide for 30 min at room temperature in the dark). On completion of the alkylation, each sample was diluted 1:4 with 20 mM HEPES (pH 8.0). To prepare the trypsin beads for addition to the sample, an appropriate volume of beads was taken from the stock and centrifuged at 2,000 x *g* for 5 min (5^{o}C). The resulting supernatant was discarded, and an equal volume of 20 mM HEPES (pH 8.0) added to create a 50% slurry. This process was repeated three times to ensure adequate conditioning of the beads. 50% trypsin bead slurry was added to each sample to a final concentration of 4% (v/v) ($\geq$ 32 TAME units) and the samples then incubated at 37^{o}C, with constant agitation, overnight.

2.7.2 Peptide desalting by solid-phase extraction

Peptide solutions (resulting from the overnight tryptic digestion – Section 2.7.1) were desalted by means of reversed-phase solid-phase extraction (utilising OASIS® HLB cartridges). The cartridges, held in a vacuum manifold (P = 5.0 inHg $\pm$ 0.5), were conditioned with LC-MS grade acetonitrile (ACN) and subsequently equilibrated with 99% H_2O (1% ACN; 0.1% TFA). Samples were loaded into individual cartridges and washed with 99% H_2O (1% ACN; 0.1% TFA) to remove residual salts. The bound peptides were then eluted with 0.5 mL 1M glycolic acid (50% ACN; 5% TFA – for phosphoproteomics) or 70% ACN (30% H_2O – for proteomics). Desalted peptide samples for phosphoproteomics analysis were then normalised to 1 mL through addition of 0.5 mL 1M glycolic acid (80% ACN; 5% TFA), and subsequently stored on ice prior to phospho-enrichment. Desalted peptide samples for proteomics analysis were vacuum dried immediately following the desalting. All solutions for use in either desalting or phospho-enrichment protocols were prepared with LC-MS grade solvents.

2.7.3 TiO_2-Metal Oxide Affinity Chromatography (MOAC) – on-line method

Following digestion and desalting, peptide samples were incubated with TiO_2 beads (25 μg TiO_2 beads$\cdot\mu$g protein^{-1}) at room temperature for 5 min, with constant agitation. The samples were then centrifuged at 500 x *g* for 30s and 0.8 mL of the remaining supernatant removed and discarded. The remaining volume was subsequently used to resuspend the beads. The resuspended beads were then loaded into Protea C18 micro-spin columns (pre-cleaned with LC-MS grade ACN) and centrifuged at 4,000 x *g* for 3 min. Any residual beads in the incubation tubes were resuspended in 0.2 mL 1M glycolic acid (80% ACN; 5% TFA) and loaded into the columns as before. The columns were then washed sequentially with 0.2 mL 1M glycolic acid (80% ACN; 5% TFA), 98% H_2O (2% ACN), and 100 mM $NH_4CH_3COO^-$ (75% H_2O; 25% ACN) – centrifuging at 4,000 x *g* for 3 min between each wash. The phosphopeptides were then transferred from the TiO_2 layer to the reversed-phase sorbent by means of a wash with 0.2 mL 5% (v/v) NH_4OH (98% H_2O; 2% ACN), and then washed twice with 0.2 mL 98% H_2O (2% ACN; 0.1% FA). The phosphopeptides were then eluted from the C18 layer with two washes of 0.1 mL 40% H_2O (60% ACN; 0.1% FA). Samples were then vacuum-dried overnight and stored at -80 ^{o}C.

2.7.4 TIO_2-MOAC – off-line method

Digested, desalted peptide samples were incubated with TiO_2 beads (25/50 μg TiO_2 beads·μg protein^{-1}) at room temperature for 5 min, with constant agitation. The samples were then centrifuged at 500 x *g* for 30s and 0.8 mL of the remaining supernatant removed and stored on ice. The remaining volume was subsequently used to resuspend the beads which were then packed into pre-cleaned Glygen empty TopTips and centrifuged at 1,500 x *g* for 3 min. The stored 0.8 mL of each sample was then washed over the packed beads (four washes of 0.2 mL), centrifuging at 1,500 x *g* for 3 min between loading of each aliquot. Each micro-spin column was then washed sequentially with 0.2 mL 1M glycolic acid (80% ACN; 5% TFA), 0.2 mL 100 mM $NH_4CH_3COO^-$ (75% H_2O; 25% ACN), and 0.1 mL 98% H_2O (2% ACN) – centrifuging at 1,500 x *g* for 3 min between each wash. The bound phosphopeptides were then eluted with four 50 μL washes with 5% (v/v) NH_4OH (98% H_2O; 2% ACN) – centrifuging at 1,500 x *g* for 2 min and collecting the eluate of each wash. The resulting phosphopeptide solution was then acidifed to 10% TFA on ice.

For the following desalting stage, C18 or C18+GC micro-spin columns were conditioned and equilibrated with two aliquots of 0.2 mL 50% H_2O (50% ACN; 0.1% FA) and 0.2 mL 99% H_2O (1% ACN; 0.1% FA), respectively – centrifuging at 1,500 x *g* for 2 min between each wash. Each sample was then removed from ice and the 0.2 mL phosphopeptide eluate loaded into the columns. Each column was then washed with 99% H_2O (1% ACN; 0.1% FA) in duplicate and the peptides eluted by means of four, 50 μL washes with 50% H_2O (50% ACN; 0.1% FA). The resulting eluates were vacuum-dried overnight and stored at -80 ^{o}C.

2.7.5 TiO_2-MOAC – final enrichment method

Each desalted peptide sample was incubated with 50 μg TiO_2 beads·μg protein^{-1} (50% slurry in 1% TFA) at room temperature for 5 min, with constant agitation. The samples were centrifuged at 500 x *g* for 30 s, 0.8 mL of the supernatant transferred to a fresh tube on ice, and the remaining 0.2 mL used to re-suspend the beads. The remaining 0.2 mL of each sample (containing the beads) was then loaded into pre-washed Glygen empty TopTips and centrifuged at 1,500 x *g* for 3 min. Any residual beads were resuspended with 0.2 mL 1M glycolic acid (80% ACN; 5% TFA) and loaded into the columns as before. The remaining 0.8 mL of each sample was then washed over the beads (four washes of 0.2 mL), centrifuging at 1,500 x *g* for 3 min between loading of each aliquot. Each micro-spin column was then washed sequentially with 0.2 mL 1M glycolic acid (80% ACN; 5% TFA), 0.2 mL 100 mM $NH_4CH_3COO^-$ (75% H_2O; 25% ACN), and 0.6 mL 90% H_2O (10% ACN – three washes of 0.2 mL) – centrifuging at 1,500 x *g* for 3 min between each wash. The bound phosphopeptides were then eluted with four 50 μL washes with 5% (v/v) NH_4OH (90% H_2O; 10% ACN) – centrifuging at 1,500 x *g* for 2 min and collecting the eluent of each wash. The resulting phosphopeptide solution was then snap-frozen on dry-ice for 15 min and vacuum dried until no solvent remained.

2.8 Liquid Chromatography-Mass Spectrometry (LC-MS)

2.8.1 Sample reconstitution and LC separations

Immediately prior to LC-MS analysis, the samples were reconstituted with 20 μL (Velos) or 14 μL (XL) of 50 fmol·μL^{-1} enolase peptide digest (dissolved in 95% H_2O; 5% ACN; 0.1% TFA), bath sonicated for 5 min at room temperature, centrifuged at 16,000 x *g* for 5 min (5 ^{o}C), and the resulting supernatant recovered for LC-MS analysis.

Peptide and phosphopeptide LC separations – unless stated otherwise – were carried out online on Dionex Ultimate 3000 nRSLC (Velos) and Waters NanoACQUITY UPLC systems (XL). Separations on the Dionex system were performed using an Acclaim® PepMap RLSC C18 Analytical Column (75 μm x 25 mm, 3 μm, 100$\AA$) and an Acclaim® PepMap 100 C18 Trap Column (100 μm x 2 cm, 5 μm, 100$\AA$). Solvent A and B were 98% H_2O (2% ACN; 0.1% FA) and 20% H_2O (80% ACN; 0.1% FA) respectively. All sample injections were equal to 3.0 μL; being loaded onto the trap column at a flow rate of 8 $\mu L \cdot min^{-1}$ for 5 min. Once loaded, the samples were eluted over an 85 min gradient from 6.3 to 43.8% solvent A. Following elution, the column was cleaned with 90% solvent B for 10 min, and subsequently equilibrated with 6.3% solvent A for 10 min.

Separations on the Waters system were performed using a Waters BEH130 (ethylene bridged hybrid) C18 analytical column (75 μm x 150 mm, 1.7 μm, 300Å) and a Waters UPLC 2G trap column (180 μm x 20 mm, 5 μm, 100 Å). Solvent A and B were 100% H_2O (0.1% FA) and 100% ACN (0.1% FA) respectively. Samples were injected at a volume of 3.0 μL and were loaded on to the trap column at a flow rate of 2 $\mu L \cdot min^{-1}$ for 8 min. Elution was performed over an 100 min gradient from 5 to 35% solvent A. Following the gradient, the column was cleaned with 90% solvent B for 10 min, and then equilibrated to 99% A; 1% B for 5 min.

2.8.2 Mass spectrometric analysis

Analyses were performed on Thermo Scientific LTQ Orbitrap-Velos and LTQ Orbitrap-XL hybrid instruments (as indicated), operated in data-dependent acquisition (DDA) mode. In the DDA method used – unless stated otherwise – a full MS^1 survey scan (m/z 350-1500) was performed at a resolution of 30,000 FWHM (at m/z 400); the ions being analysed in the Orbitrap. The top 7 (Velos) or top 5 (XL) most intense multiply charged precursor ions present in the MS^1 scan were automatically mass-selected and fragmented by collision-induced dissociation (CID – normalised collision energy = 35%) with multi-stage activation enabled, and analysed in the LTQ-Velos/XL linear ion trap (m/z 190-2000). Neutral losses of 98, 49, 32.7, and 24.5 were accounted for and dynamic exclusion was enabled (avoiding repeat analysis of identical precursor ions within a 60 s window).

2.8.3 Thermo LTQ-Orbitrap-Velos and -XL mass spectrometers

The Thermo LTQ-Orbitrap-Velos and -XL instruments both consist of a combination of orbitrap and linear quadrupole ion trap mass analysers (Fig 2.1); however, the instruments differ in the composition of the ion optics prior to the ion trap. The overall schematics of both mass spectrometers are shown in Fig 2.1. Peptide ions, separated on the nano-LC system connected online to the mass spectrometer, are introduced to the instrument via the electrospray ionisation process (Chapter 1; Fig 1.5). The ions are then focused by the ion optics and guided into the linear ion trap.

The XL ion optics are based on a high-pressure, skimmer-based design, whereas the Velos employs a stacked-ring ion guide (S-lens). The S-lens greatly improves the efficiency of the transfer of ions from the ion source to the ion trap, thus improving the sensitivity and limit of quantification (LOQ) of the instrument.

Once in the ion trap, the ions are trapped radially and axially through the application of an electrical field to the hyperbolic rod quadrupoles that constitute the ion trap. This creates a potential well, where the ions are unable to exit the trap. The ions, now confined within the trap, proceed to orbit at a frequency directly related to their *m/z*, as described by the Mathieu equation. Once collected, a packet of the ions is ejected and passed through to the C-trap, where they are then

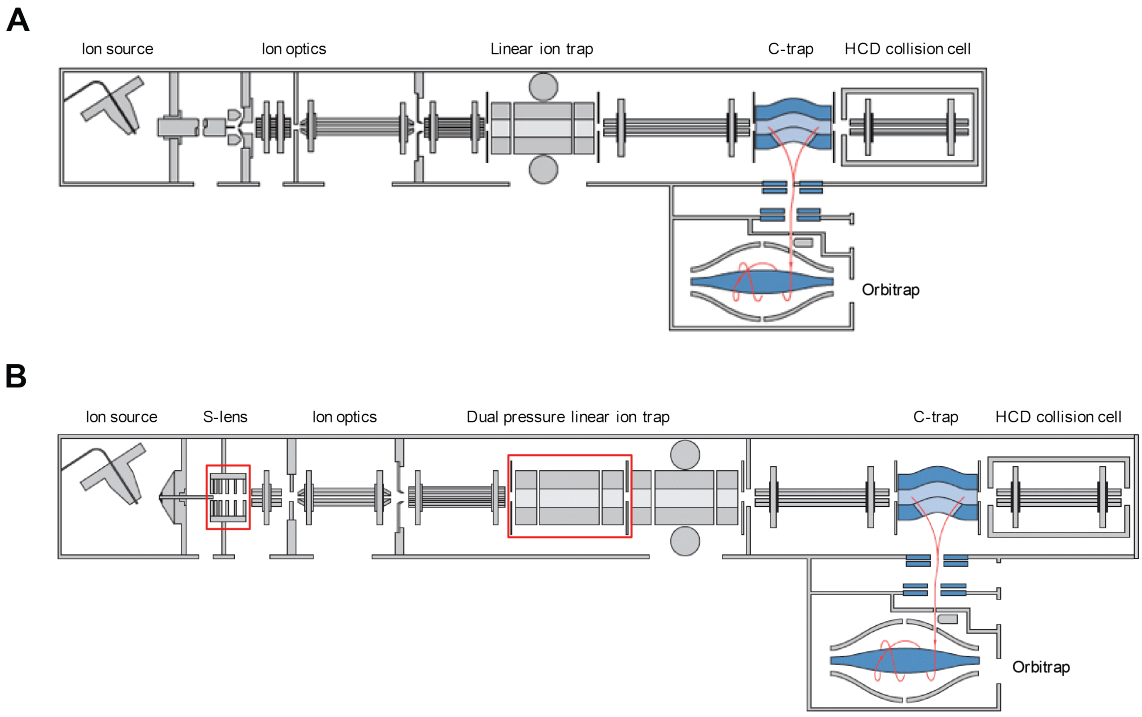


Figure 2.1: Schematics of the orbitrap-XL and -Velos mass spectrometers. The major components of the orbitrap-XL (A) and orbitrap-Velos (B) are shown (not to scale). The main differences between the instruments are highlighted with red boxes. Diagrams adapted from: planetorbitrap.com

tangentially injected into the orbitrap mass analyser for high resolution and high mass accuracy MS^1 analysis. Fig 2.2A shows an example MS^2 spectrum produced in an ion trap.

The orbitrap consists of two electrodes: the outer, barrel-like electrode and the inner, spindle-like electrode. Due to the RF and DC fields applied to the electrodes, ions orbit around the inner electrode on helical trajectories at a frequency determined by the ions' *m/z* ratios. The mass spectrum of the ions present within the orbitrap is obtained from a Fourier transform of the image current produced from the trapped ions. Fig 2.2B shows an example MS^1 spectrum measured in an orbitrap.

In DDA methods, the *n* most intense, multiply charged ions (i.e. those likely to represent tryptic peptides) are then sequentially selected by the ion trap, fragmented, and the *m/z* ratios of the resulting fragment ions measured by electron multipliers adjacent to the ion trap. This process is referred to as MS^2 analysis and the overall time taken to measure one MS^1 scan and *n* x MS^2 scans is referred to as a duty cycle. The duty cycle of an instrument is related to its sensitivity (i.e. how quickly/efficiently it can fill the ion trap), scan speed of both analysers, the number of ions chosen to be selected for MS^2 (*n*), and the fragmentation method used. Alongside the impact on the number of peptide identifications can be made, the duty cycle also directly impacts the quality of MS^1-based quantitative data, as a shorter duty cycle allows for more quantitative (MS^1) data-points per peak, and thus a more accurate peak shape. In this respect, a distinct advantage of the Velos instrument is that, due to its increased sensitivity, its duty cycle is much shorter than that of the XL and thus a greater number of peptides can be identified in the same chromatographic separation time. Moreover, this allows for shorter overall separation times (therefore increasing throughput), without sacrificing the number of identifications and quantitative data quality.

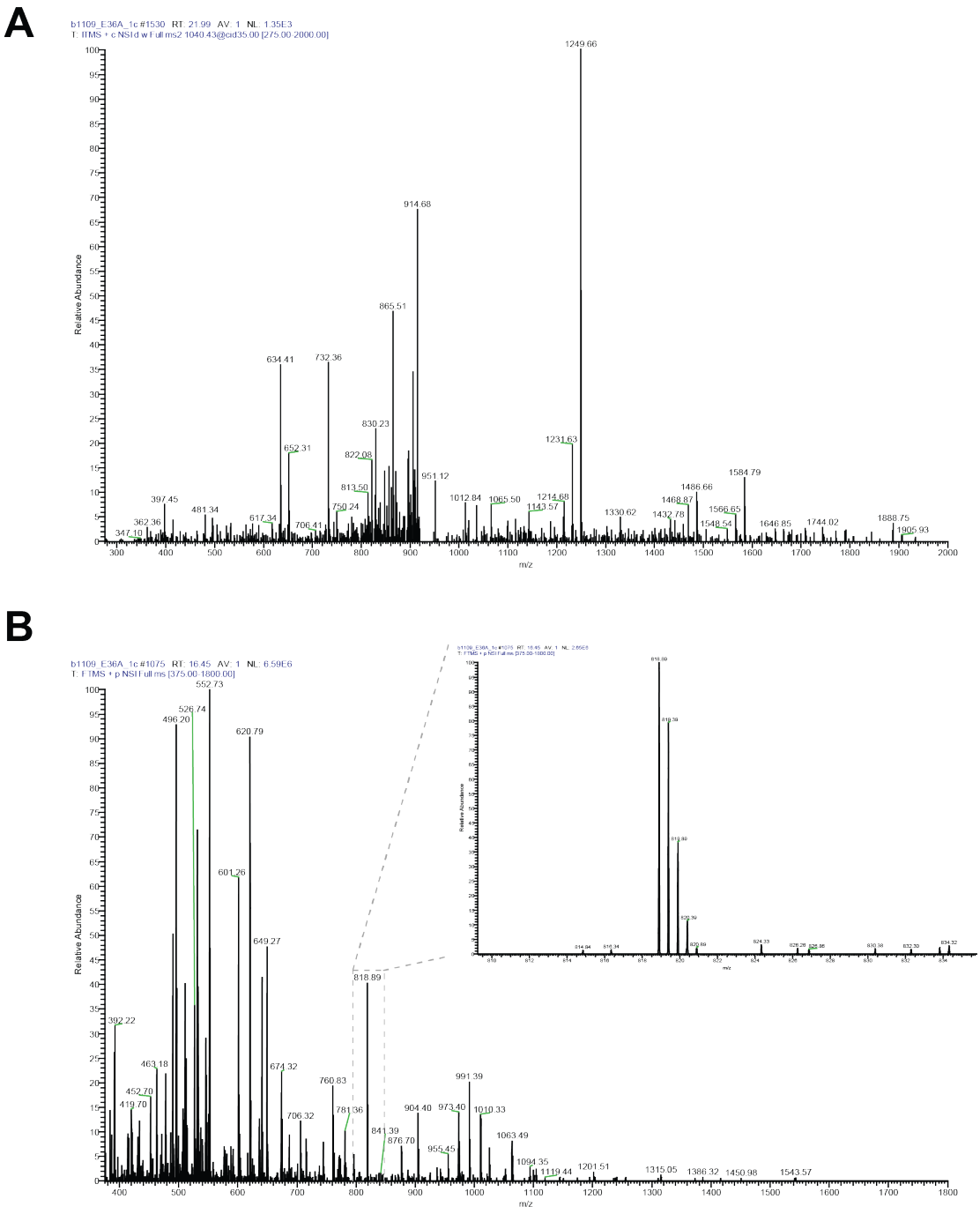


Figure 2.2: Example spectra produced in ion trap and orbitrap mass analysers. (A) An MS^2 spectrum produced in an ion trap instrument. (B) An MS^1 spectrum produced in an orbitrap instrument. The zoomed area shows the isolated spectrum of a specimen precursor ion, demonstrating the resolution of the orbitrap mass analysis (at 30,000 FWHM).

2.9 Data analysis

2.9.1 Mascot database searches

Mascot Daemon and Distiller (v2.3.0.0 and v2.4.2.0 respectively – Matrix Science, London, UK) were used in conjuction to automate the conversion of Thermo Scientific .raw files to MS^2 smoothed and centroided peak lists (.mgf files), and to search the peak lists against the Uniprot/Swissprot human database (downloaded on 03/10/2012). LTQ Orbitrap-Velos and-XL data were searched using the following criteria: $\pm$10 ppm precursor and $\pm$600 mmu fragment ion m/z tolerances; enyzme = trypsin (2 missed cleavages tolerated); fixed modification: carbamidomethyl (C); variable modifications: gln$\rightarrow$pyro-glu (N-term Q), oxidation (M), phospho (ST), and phospho (Y).

Phosphopeptide identification data produced by Mascot search engine results were collated using a combination of a Perl script and Peptide ANalysis and Database Assembly (PANDA) software (v1.1) – both developed in-house. The data were algorithmically curated to only include unique phosphopeptide ions (accurate to $\pm$10 ppm) with a *q*-value $\leq$ 0.05 (calculated via comparison to searches against a randomised database). All phosphopeptides with a Mascot delta score $<$ 10 were reported as "Protein (residues); charge; modification(s)". All those with a score $\geq$ 10 were reported as the specific phosphorylation site (e.g. MAPK3 p-T202 p-Y204). Protein identity was inferred from peptides identified from the aforementioned Mascot search engine results (excluding phospho (ST) and (Y) variable modifications). Identified proteins were required to meet the following thresholds to be included in further analyses: minimum peptide score $\geq$ 20; $\geq$ 2 unique peptides; minimum protein score $\geq$ 60.

2.9.2 Phosphopeptide quantification

Phosphopeptides and peptides were quantified using the in-house developed PESCAL (Peak Statistic Calculator) software (written in Python v2.7); automatically generating XICs for the first three isotopes of each peptide ion within the created database ($\pm$7 ppm m/z tolerance; $\pm$1.5 min retention time tolerance; isotope correlation $>$ 0.8), subsequently calculating the peak heights of each constructed XIC. A specimen XIC is presented in Fig 2.3. Data were then processed further and analysed within Microsoft Excel (2007/2010), R (v3.0.0 – *reshape2*, *ggplot2*, *plot3D*, *XMLRPC*, *RCytoscape*, *igraph*, *limma*, and *gplots* packages)[255–263] and Cytoscape (v.2.8.3)[264]. Proteomics data were processed as above; however, the abundance of each protein was inferred by computing the sum of the intensities of all the peptides in the database corresponding to the relevant protein. Where applicable, stoichiometry ratios were calculated by subtracting the $\log_2$ fold-ratio (vs control) of the total protein from the $\log_2$ fold-ratio (vs control) of the individual phosphopeptide. Phosphopeptides not possessing protein data in this context were removed from downstream analyses.

2.9.3 Quantitative data: pre-processing and differential abundance analysis

Phosphopeptide ion intensities were $\log_2$ transformed, resulting in intensities symmetrically distributed about the mean (approximating normal distributions). The values were then between-sample quantile normalised to produce statistically identical distributions across the sample array[265].

Differential abundance analysis was conducted using the *limma* package within the R statistical computing environment. This involved the fitting of a linear model to the quantitative data for each phosphopeptide ion within the dataset (Fig 2.4; equation 2.1). Briefly, a design matrix (X –

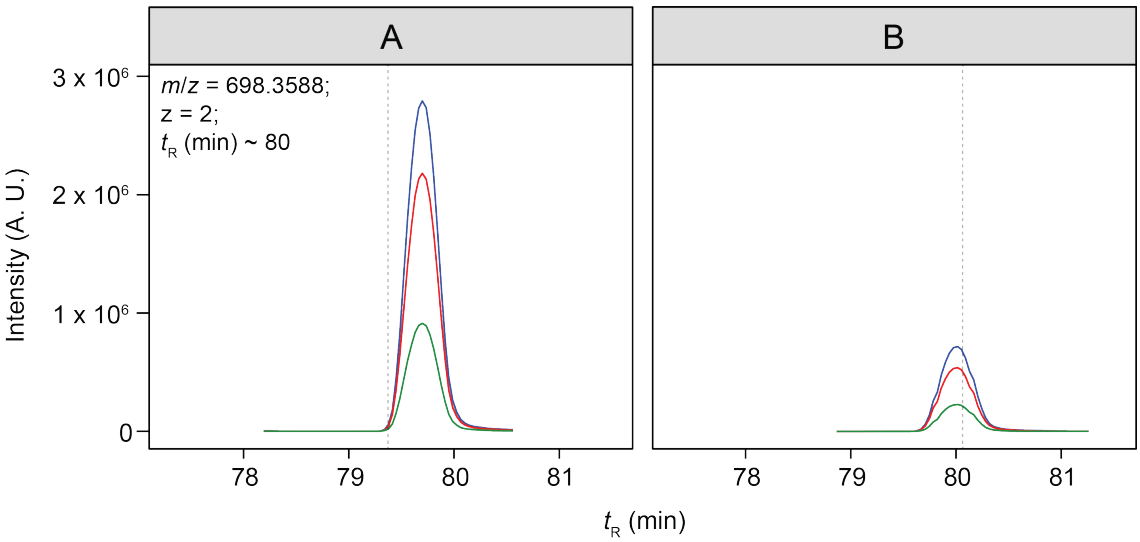


Figure 2.3: A specimen XIC produced by PESCAL. (A) An XIC representing an ion of m/z = 698.3588 and $t_R \approx 80$ min under control conditions. (B) An XIC representing the same ion as in (A), under drug treated conditions. Blue line, 1st isotope; red line, 2nd isotope; green line, third isotope. Dashed, grey line represents the predicted retention time in each case.

specific to each experiment) was created, which served to define the linear relationships between individual treatments and observed values (phosphopeptide peak heights) – Fig 2.4; equation 2.2. A linear model was then fitted to each phosphopeptide ion in the dataset, and the coefficients (α) calculated – this approximately representing the arithmetic mean of technical and/or biological replicates for each condition within the experiment design. The biological comparisons of interest were then specified in a contrast matrix (β), which described the linear relationships between the individual coefficients previously calculated (Fig 2.4; equations 2.3 and 2.4). This calculation produced $\log_2$ ratios for each of the comparisons specified ($log_2(drug) - log_2(control)$, for example). An overall, generalised form of the linear model equation summarising the aforementioned processes is shown in Fig 2.4; equation 2.5. Statistical significance of differential expression was then determined by means of empirical Bayes shrinkage of standard deviations[262,266]. The generated *P*-values were then adjusted for multiple testing using the Benjamini-Hochberg procedure to control the type I error rate[267].

$$E[y_j] = X.\alpha_j \tag{2.1}$$

$X =$

Sample	Drug	Control
drug_rep$_1$	1	0
drug_rep$_2$	1	0
control_rep$_1$	0	1
control_rep$_2$	0	1

$$= \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \tag{2.2}$$

$C^T =$

Sample	Drug – Control
drug_rep$_1$	1
drug_rep$_2$	1
control_rep$_1$	-1
control_rep$_2$	-1

$$= \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \tag{2.3}$$

$$\beta_j = C^T.\alpha_j \tag{2.4}$$

$$E[y_i] = X.\beta_j + \epsilon \tag{2.5}$$

Figure 2.4: Linear models for differential expression analysis. Equation 2.1: the linear model fitted to each phosphopeptide ion, where y_j = the quantitative data for the phosphopeptide ion *j*, X = the specified design matrix, and α_j = the vector of coefficients to be calculated. Equation 2.2: constructing the design matrix specific to the experiment. In this case, the replicates representing drug and control treated samples are assigned 1 or 0 according to which coefficient (α) they represent (*Drug* or *Control*). Equation 2.3: defining the biological comparisons to be computed. C^T = the contrast matrix representing the linear relationships between the drug and control treated samples in order to produce the contrast of interest: *Drug – Control*. Equation 2.4: β_j = the contrasts to be calculated, C^T = the contrast matrix, and α_j = the vector of previously calculated coefficients. Equation 2.5: the summarised, general equation for the linear models fitted to the data; y_j = the quantitative data for the phosphopeptide ion *j*, X = the specified design matrix, β_j = contrasts to be calculated, and ϵ = the error associated with the model.

Chapter 3

Enhancement of “single-shot” phospho-enrichment and mass spectrometric methodologies to maximise phosphoproteome coverage

3.1 Introduction and aims of the study

The enrichment of phosphopeptides from complex biological matrices is key to the success of phosphoproteomics experiments. As outlined in Section 1.6, this is necessary as a result of the substoichiometric levels of phosphorylation sites in biological systems. As such, enrichment methods bring phosphorylated peptides within the dynamic range of modern mass spectrometers. TiO_2-based enrichment methods have been developed extensively in recent years and have proved to be superior to Fe^{3+}-IMAC and antibody-based approaches with respect to the enrichment of pSer and pThr peptides in terms of a number of factors, including: the number of phosphopeptides detected, the specificity of the enrichment, and the simplicity and speed of the method[190,237,239,268]. Due to their relative simplicity, TiO_2-based methods are particularly well suited to label-free phosphoproteomics analyses. This is because, as large two-dimensional separations are generally incompatible with label-free analyses, "single-shot" enrichments (i.e. one enrichment per sample) are necessary. As a result, such single-shot enrichments must be optimised to be as specific and sensitive as possible in order to provide maximal coverage of the phosphoproteome from a single extraction.

The utility of single-shot enrichment methods, both in terms of qualitative and quantitative data has been previously demonstrated[244]. Work from the Cutillas lab has also characterised a single-step TiO_2-based enrichment method (Fig 3.1A), primarily for label-free analyses, that has been shown to be both quantitatively reproducible and accurate (mean CV post-normalisation = 23.8% and error $\leq$ 30% for the majority of peptides), alongside being qualitatively efficacious (approximately 500 phosphopeptides per sample at $\sim$90% enrichment)[141]. Despite this method's utility, a number of issues were present with the protocol; namely, a lack of reversed-phase mechanical robustness, relatively low recovery compared to two-dimensional methods, and most pertinently, *N*-formylation of phosphopeptides during vacuum drying (Fig 3.1B). *N*-formylation of amino groups present within the phosphopeptides occurred in this context due to the heating of the sample in the presence of 10% formic acid – utilised to neutralise the NH_4OH necessary to elute the phosphopeptides from the TiO_2 – while drying. This unexpected modification resulting in a lower number of phosphopeptide identifications (as this modification was not included in the database searches), whilst concomitantly convoluting the quantitative analysis, as the rate of *N*-formylation for each peptide is unknown, resulting in unknown ratios of *N*-formylated vs unmodified phosphopeptide ions.

The aim of the experiments presented in this Chapter was to optimise and enhance the original phospho-enrichment method as presented in Montoya *et al.*[141]; in terms of absolute recovery of phosphopeptides, a reduction in the *N*-formylation rate of said peptides, and an improvement of the protocol's robustness and efficiency. In order to achieve this, several iterations of the phospho-enrichment method were devised and tested to provide a single, efficient method that would provide adequate depth of analysis for future experiments. Furthermore, mass spectrometric parameters, such as the precursor ion selection and fragmentation methods used to sequence the recovered phosphopeptides, were also tested.

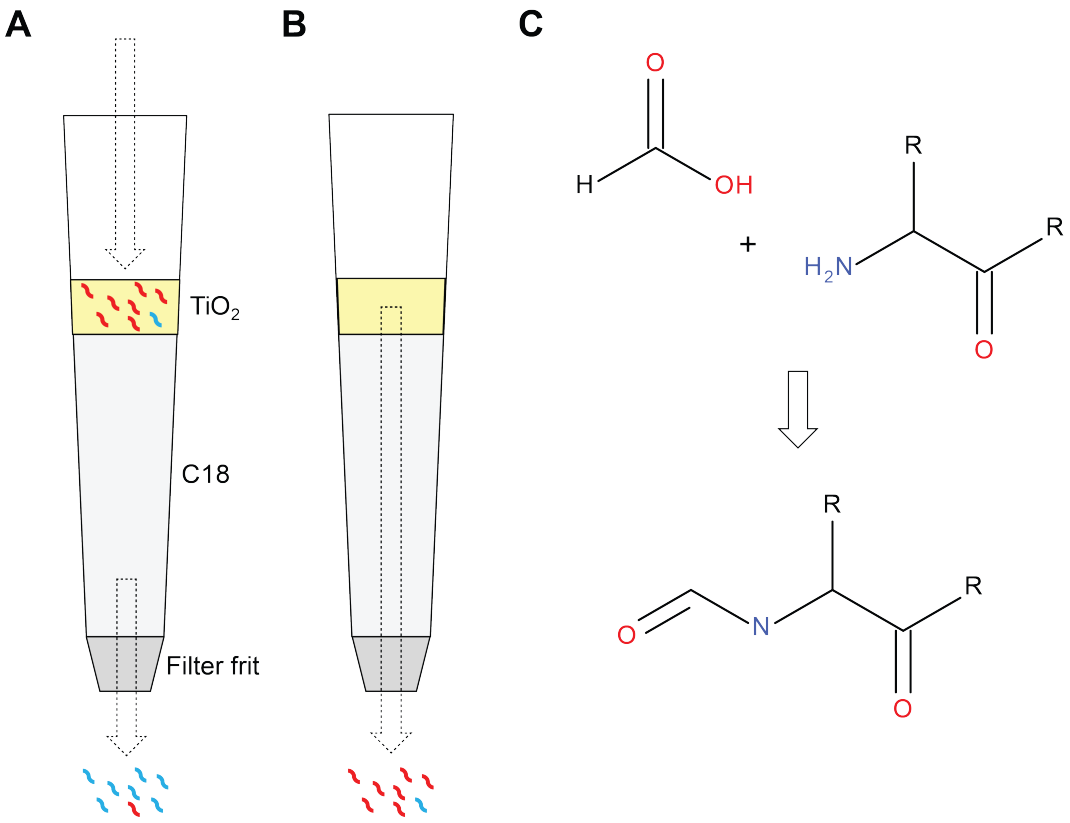


Figure 3.1: The original phospho-enrichment methodology as presented in Montoya *et al.* (A) Following desalting, the purified peptide mixture is mixed and incubated with TiO_2 beads. This mixture is then loaded onto the C18 sorbent (packed in a micro-spin column format) in the presence of a high percentage of ACN (to prevent C18 binding), 5% TFA and 1M glycolic acid (to prevent non-phosphopeptide binding to the TiO_2). (B) Following washing steps to improve specificity of phosphopeptide binding, the phosphopeptides are eluted with 5% NH_4OH in the presence of a high percentage of ACN. The resulting phosphopeptide mixture is acidified with formic acid (10% final concentration) and vacuum-dried. Blue, non-phosphopeptides; red, phosphopeptides. (C) In the presence of formic acid and vacuum-drying at relatively high temperatures (30-40 ^{o}C), free amino groups (e.g. *N*-termini, Arg, and Lys residues) can become formylated, adding 31 Da to the peptide.

3.2 Optimisation of phospho-enrichment conditions

3.2.1 Effects of on-line desalting on phosphopeptide recovery

In order to circumvent the issue of *N*-formylation and therefore improve phosphopeptide recovery, it was hypothesised that an on-line enrichment method may be more effective. Such a method would involve the elution of the phosphopeptides from the TiO_2 layer and immediate recaptured on a reversed-phase (C18) layer coupled on-line in a micro-spin column (Fig 3.2A-B). The C18 layer could then be washed to remove the remaining NH_4OH, formic acid, and ammonium formate salt (Fig 3.2B), thereby removing the acidification step. The remaining peptides could then easily be eluted from the C18 layer with a high concentration of ACN (Fig. 3.2C). To test this hypothesis, an experiment was designed to compare the two enrichment methods. To achieve this, a pooled MCF7 lysate was separated into eight 500 μg aliquots. Following reduction, alkylation, and an overnight tryptic digestion (at 37 ^{o}C), these eight samples were desalted via solid-phase extraction using OASIS® HLB cartridges (1 cc; 30 mg) and the phosphopeptides extracted via TiO_2-MOAC – four extractions utilising the published method[141] and four using the on-line method (Fig 3.2A-C). Once vacuum-dried, these extractions were analysed on a Thermo LTQ-Orbitrap-XL instrument after having been separated on a 45 min linear gradient (5-35% ACN) on a Waters nanoACQUITY UPLC system. In each case the ions were fragmented by CID-MSA (top 5). The results of this comparison are shown in Fig 3.2D.

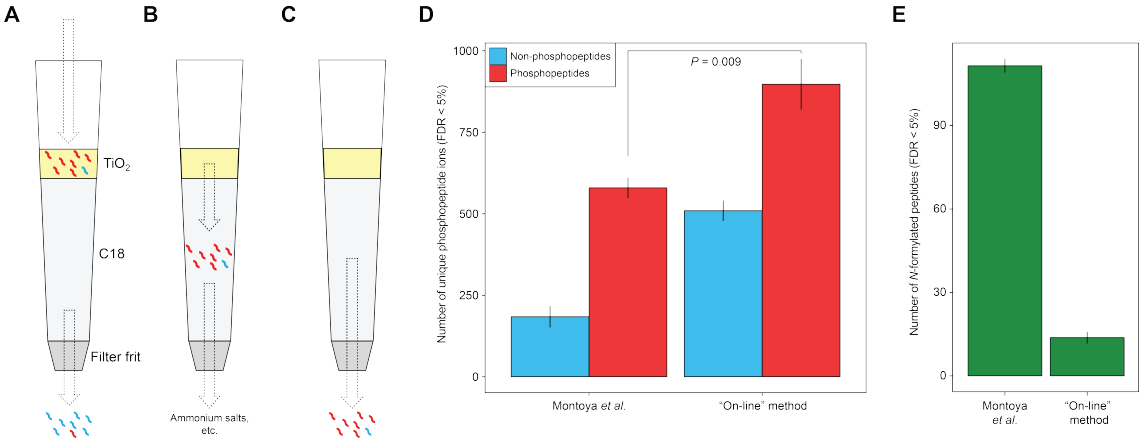


Figure 3.2: An "on-line" desalting phospho-enrichment methodology improves phosphopeptide recovery (A) Following desalting, the purified peptide mixture is mixed and incubated with TiO_2 beads. This mixture is then loaded onto the C18 sorbent in the presence of a high percentage of ACN (to prevent C18 binding), 5% TFA and 1M glycolic acid (to prevent non-phosphopeptide binding to the TiO_2). (B) Following washing steps, the phosphopeptides are eluted with 5% NH_4OH in the presence of a low percentage of ACN. The phosphopeptides are therefore removed from the TiO_2, but retained by the C18 layer. The C18 is then washed with a low percentage of ACN containing 1% formic acid in order to neutralise any remaining NH_4OH. (C) The remaining phosphopeptide ions are then eluted from the C18 layer with a high percentage of ACN and vacuum-dried for further analysis. (D) Qualitative comparison of the Montoya *et al.* method and the on-line method. (E) Comparison of the number of *N*-formylated peptides identified using both methods. Error bars represent SEM (n = 4). *P*-values were determined via a two-tailed Student's *t*-test.

As Fig 3.2D demonstrates, and as anticipated, the on-line method was able to substantially increase the recovery of unique phosphopeptide ions from single-step TiO_2 enrichments (580 vs 898; P = 0.009). Moreover, the number of *N*-formylated peptides was significantly reduced (Fig 3.2D – 112 vs 14; P = 9.2 x 10^{-8}). Despite these clear improvements, the efficiency of enrichment, defined as the ratio between identified phosphopeptides and non-phosphopeptides, was significantly poorer (76.0% vs 63.6%; P = 0.029).

In an effort to improve the enrichment efficiency, an experiment was designed to compare the newer on-line method with the same method with an additional 200 μL 100 mM ammonium acetate (25% ACN) wash prior to transfer of the phosphopeptides from the TiO_2 layer to the

C18. It was hypothesised that, as high salt concentrations aid peptide binding to C18[269] and that ammonium acetate has previously been shown to improve the specificity of phosphopeptide binding to TiO_2 [141], this extra wash would improve both the yield and specificity of the enrichment. As with the data shown in Fig 3.2, this experiment involved the comparison of four individual TiO_2 extractions (from 500 μg digests derived from a pooled MCF7 lysate) for each method. The data presented in Fig 3.3 demonstrate that the extra salt wash did not affect the number of phosphopeptides recovered and did not have a significant impact on the enrichment efficiency (70.3% and 70.6% first iteration and extra wash, respectively).

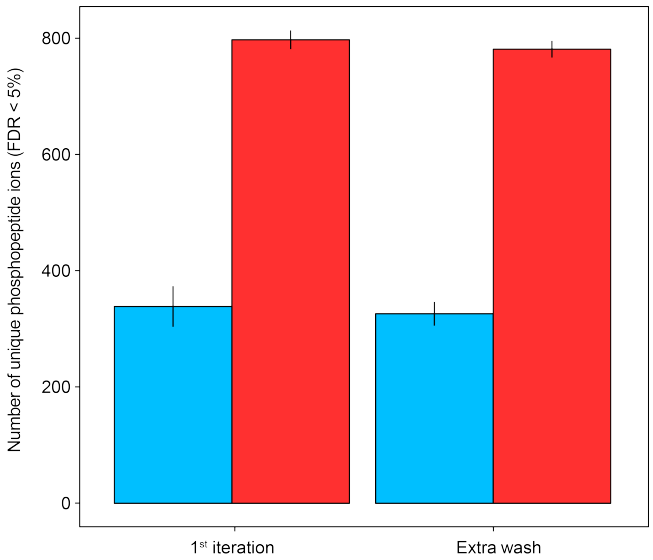


Figure 3.3: An additional salt wash does not improve phosphopeptide enrichment. Qualitative comparison of the first iteration of the "on-line" method as in Fig 3.2, and the identical method with an additional 100 mM ammonium acetate (25% ACN) wash prior to TiO_2 elution. Error bars represent SEM (n = 4). Blue, non-phosphopeptides; red, phosphopeptides.

3.2.2 Effects of "off-line" desalting

Despite the effectiveness of the on-line method, the specificity of enrichment was poorer than the original method. The source of this lack of specificity was posited to be due to high retention of non-phosphorylated peptides on the C18 layer. To combat this, it was hypothesised that, concurrent with the unforeseen availability of empty micro-spin columns that could be packed in-house, the enrichment and desalting stages could be separated. This presented several advantages. Firstly, unwanted (non-phosphopeptides) would not have the opportunity to bind to the reversed-phase and thus contaminate the enrichment. Secondly, the volume remaining following the incubation with the TiO_2 beads could be washed over the packed TiO_2 to encourage further phosphopeptide binding. Finally, the TiO_2 could be washed multiple times with high pH NH_4OH elution buffer, as concerns over the pH stability of the C18 layer were rendered irrelevant. This latter factor was of particular importance, as several groups have noted the efficacy of eluting off the TiO_2 multiple times[141,240]. A schematic of this off-line methodology is shown in Fig 3.4A-D.

To further investigate conditions that might enhance phosphopeptide recovery, two different reversed-phase sorbents were compared: C18 and C18 combined with graphitic carbon (GC). Graphitic carbon has been explored as an alternative to C18 due to its superior chemical and physical stability (particularly with regards to pH extremes). GC retains compounds via a mixed-mode mechanism, involving both charge-dependent/ion-exchange alongside C18-like hydrophobic interactions[270]. It was reasoned that, as phosphopeptides are generally more hydrophilic, this sorbent (in combination with C18) may improve phosphopeptide retention in the desalting stage. To test these two hypotheses, twelve independent TiO_2 independent extractions (from 500 μg digests derived from a pooled MCF7 lysate) were performed. Four extractions were performed using the on-line methodology, four using the off-line method with C18 only, and four using the off-line method with C18 and GC. In order to prevent *N*-formylation in these samples between enrichment and desalting, the samples were acidifed with 10% TFA – this also acting as an ion-pairing agent in the subsequent desalting stage, therefore aiding the binding of hydrophilic peptides to the reversed-phase – and were kept at 0 ^{o}C. The results of this experiment are presented in Fig 3.4E.

The results shown in Fig 3.4E show that the off-line methodology drastically improved the specificity of the enrichment (85%, 96%, and 95% for the on-line, off-line (C18), and off-line (C18+GC) methods respectively). More importantly, as predicted, the C18+GC reversed-phase sorbent facilitated the recovery of a significantly greater number (approximately 1.6-fold) of unique phosphopeptide ions than either the on-line or C18-based off-line method, whilst maintaining specificity. As a result, this method was taken forward for further development.

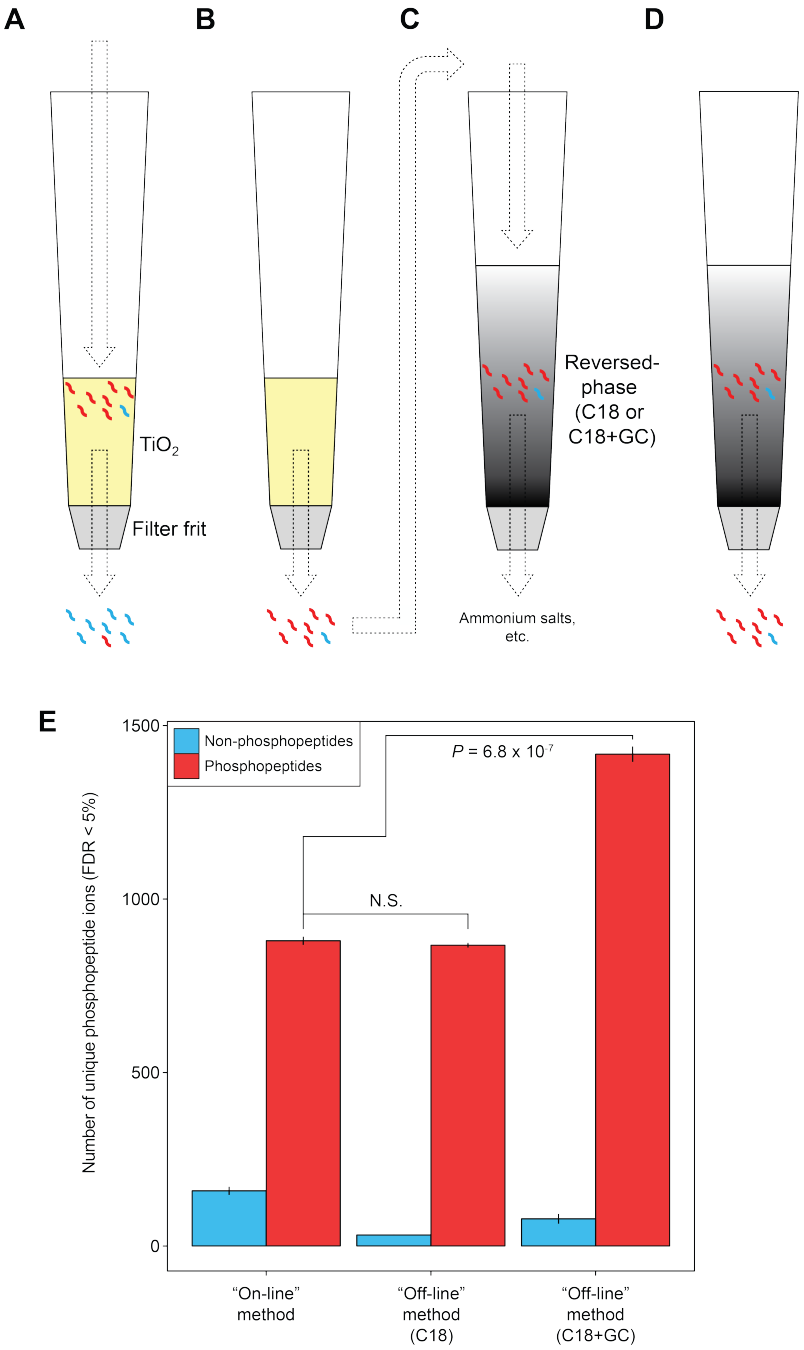


Figure 3.4: Decoupling enrichment and desalting improves phosphopeptide recovery. Qualitative comparison of the "on-line" method as in Fig 3.3, and the newly devised 'off-line" method. (A) Phosphopeptides are incubated with TiO_2 beads and loaded into empty micro-spin columns. (B) The remaining phosphopeptides are eluted from the TiO_2 layer using 5% NH_4OH, acidified (to a final concentration 10% TFA), and kept at 0 ^{o}C. (C-D) Post-acidification, the phosphopeptides are loaded into a conditioned reversed-phase-containing micro-spin column and desalted (to remove the remaining TFA and ammonium salts thereof). Error bars represent SEM (n = 4). P-values were determined via a two-tailed Student's t-test.

3.2.3 Optimisation of off-line desalting conditions

Having demonstrated that the off-line desalting method, combined with the C18-+GC sorbent, improved the yield of phosphopeptides from single extractions, a number of further enrichment conditions were adjusted in an effort to improve recovery. To this end, the amount of TiO_2 added to the reaction, C18+GC elution conditions, and TiO_2 incubation conditions were tested. The TiO_2 and elution conditions were tested at two different concentrations: 25 and 50 μg TiO_2 · μg protein^{-1} (Fig 3.5A) and 50 and 60% ACN (Fig 3.5B) – both of these comparisons were performed in quadruplicate. In order to test the loading/incubation conditions for the TiO_2, four different buffers were used: 1M glycolic acid (80% ACN; 5% TFA), 1M glycolic acid (80% ACN; 0%TFA), 2M glycolic acid (80% ACN; 0%TFA), and 2M lactic acid (80% ACN; 0%TFA), as proposed by Kettenbach and Gerber[240] – these comparisons were performed in triplicate.

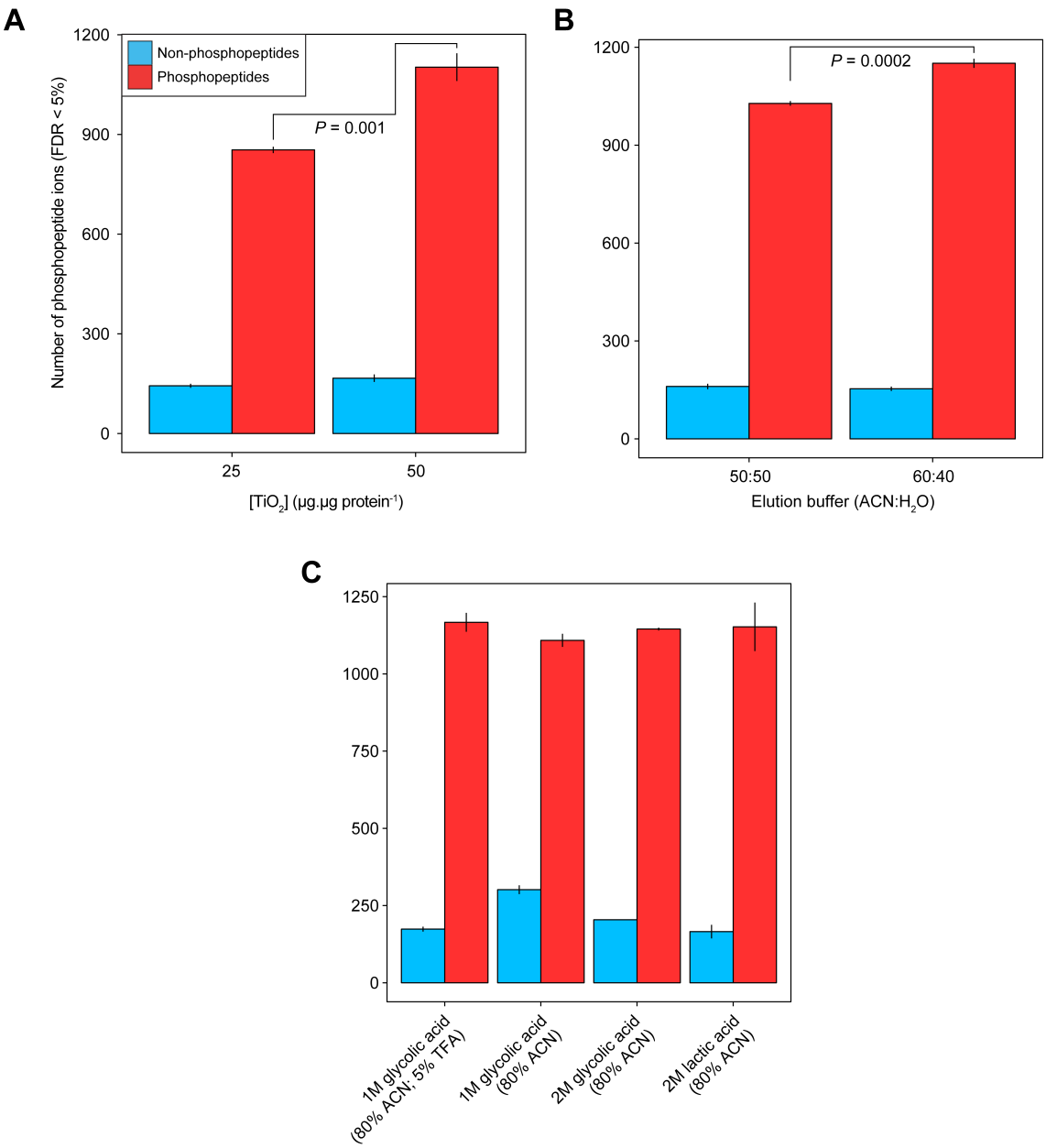


Figure 3.5: Optimisation of off-line desalting conditions for maximal phosphopeptide recovery. (A) Optimisation of TiO_2 concentration (n = 4). (B) Optimisation of ACN concentration for the elution of phosphopeptides from the C18+GC layer (n = 4). (C) Optimisation of TiO_2 loading buffer conditions (n = 3). Error bars represent SEM. P-values were determined via a two-tailed Student's t-test.

The results presented in Fig 3.5A demonstrate that a doubling of the concentration of TiO_2

present in the initial incubation did not result in a doubling of the number of identified phosphopeptides; however, the number of recovered phosphopeptides did increase significantly (853 vs 1102; $P = 0.001$). Moreover, this increase did not correspond with a significant increase in non-phosphopeptides (144 vs 167; $P = 0.124$), therefore resulting in a slight increase in specificity. As with the increase in TiO_2 concentration, increasing the proportion of ACN in the C18+GC elution buffer also significantly increased the number of phosphopeptides recovered (Fig 3.5B – 1028 vs 1151; $P = 0.0002$). Fig 3.5C shows the results from the comparison between TiO_2 loading buffers. These data demonstrate that there was no significant difference between the buffers in terms of absolute phosphopeptide recovery; however, the number of non-phosphopeptides, and therefore the specificity of enrichment, was poorer for those samples enriched in 1M glycolic acid without the presence of TFA. The specificity of enrichment was restored however, when a higher concentration of glycolic acid or lactic acid were used. Alongside these findings, it was observed that the variability in phosphopeptide recovery was higher for those samples enriched in the presence of 2M lactic acid.

3.2.4 Effects of removing phosphopeptide acidification

As previously discussed, the purpose of the acidification of the phosphopeptides is to neutralise the NH_4OH used to elute them from the TiO_2 beads, so as to prevent the loss of phosphate groups through β-elimination (i.e. base hydrolysis). Following discussions with Dr N. Morrice (University of Dundee)[271] however, it was postulated that the rate of β-elimination of phosphopeptides is only significant in the presence of high pH ($>$ 12) and a strong base (e.g. $Ba(OH)_2$ [$pK_b \approx$ -2.0], compared to $pK_b \approx 4.8$ for NH_4OH), and at temperatures greater than those experienced by the phosphopeptides at any stage during the enrichment process. Moreover, snap-freezing (on $CO_{2(s)}$) prior to vacuum drying, further prevents any potential base hydrolysis of phosphate groups and encourages sublimation of the volatile ammonium compound.

As a result of these discussions, it was hypothesised that the acidification step, and therefore the desalting of the peptides via C18+GC, was both unnecessary and may result in the loss of acidic phosphopeptides. This loss of acidic phosphopeptides occurs as, at low pH, acidic groups shift towards a state of higher protonation and are therefore less hydrophilic[271]. Alongside the loss of acidic phosphopeptides, the desalting step may result in loss of peptides both from increased contact of the peptides with plastics, and from those that bind and cannot be eluted from the C18+GC layer. To explore this hypothesis, an experiment was designed to compare the off-line desalt method (as in Fig 3.4) with one devoid of acidification and the consequential desalting step. To this end, triplicate 500 μg digests prepared from a single pooled MCF7 lysate were either enriched as per the off-line desalt method, or by means of an identical TiO_2 enrichment immediately followed by vacuum drying, with or without snap-freezing (Fig 3.6A-B). The results of this comparison are shown in Fig 3.6C.

Fig 3.6C demonstrates that the removal of the acidification steps in the protocol had no detectable effect on the absolute number of phosphopeptides recovered; however, the data showed a trend towards a decrease in phosphopeptide identifications when the samples were not snap-frozen immediately prior to vacuum-drying (1636 vs 1541, snap-frozen vs non-snap-frozen; P = 0.075). Despite no discernible difference in phosphopeptide identifications, the size of the peptide pellet recovered post-vacuum drying was much larger for those samples that were not acidified compared to those processed using off-line desalt method (data not shown). Moreover, a comparison of the chromatograms produced from off-line and no acidification (with snap-freezing) method demonstrated that, although both samples were similarly complex in the hydrophilic region (5-30 min), the latter showed much more complexity in the hydrophobic region of the chromatogram (Fig 3.6D, red box). To further investigate the effect of the off-line and new method (no acidification with snap-freezing), two phosphopeptide databases were created; one representing the combined database from the samples processed with the off-line and one representing the combined database for the samples processed with the new method. The peptide length, number of ionisable residues (to estimate hydrophilicity), and peptide charge were then computed and the number of common peptide identifications compared – this analysis is shown in Fig 3.7.

The analysis presented in Fig 3.7 demonstrates that removal of the acidification and desalting steps in the phospho-enrichment resulted in the recovery of significantly longer phosphopeptides (Fig 3.7A) and those that contained a greater proportion of ionisable residues (Fig 3.7B). The quantile-quantile plots presented in Fig 3.7C-D provide further evidence to suggest that these data are indeed from different distributions (as they do not conform to $y = x$ [red line]). Furthermore, the extent of overlap in peptide identifications between the two methods was approximately 43%, suggesting that they enrich for subtly different populations of phosphopeptides (Fig 3.7E). In concordance with these observations and as hypothesised, the removal of acidification

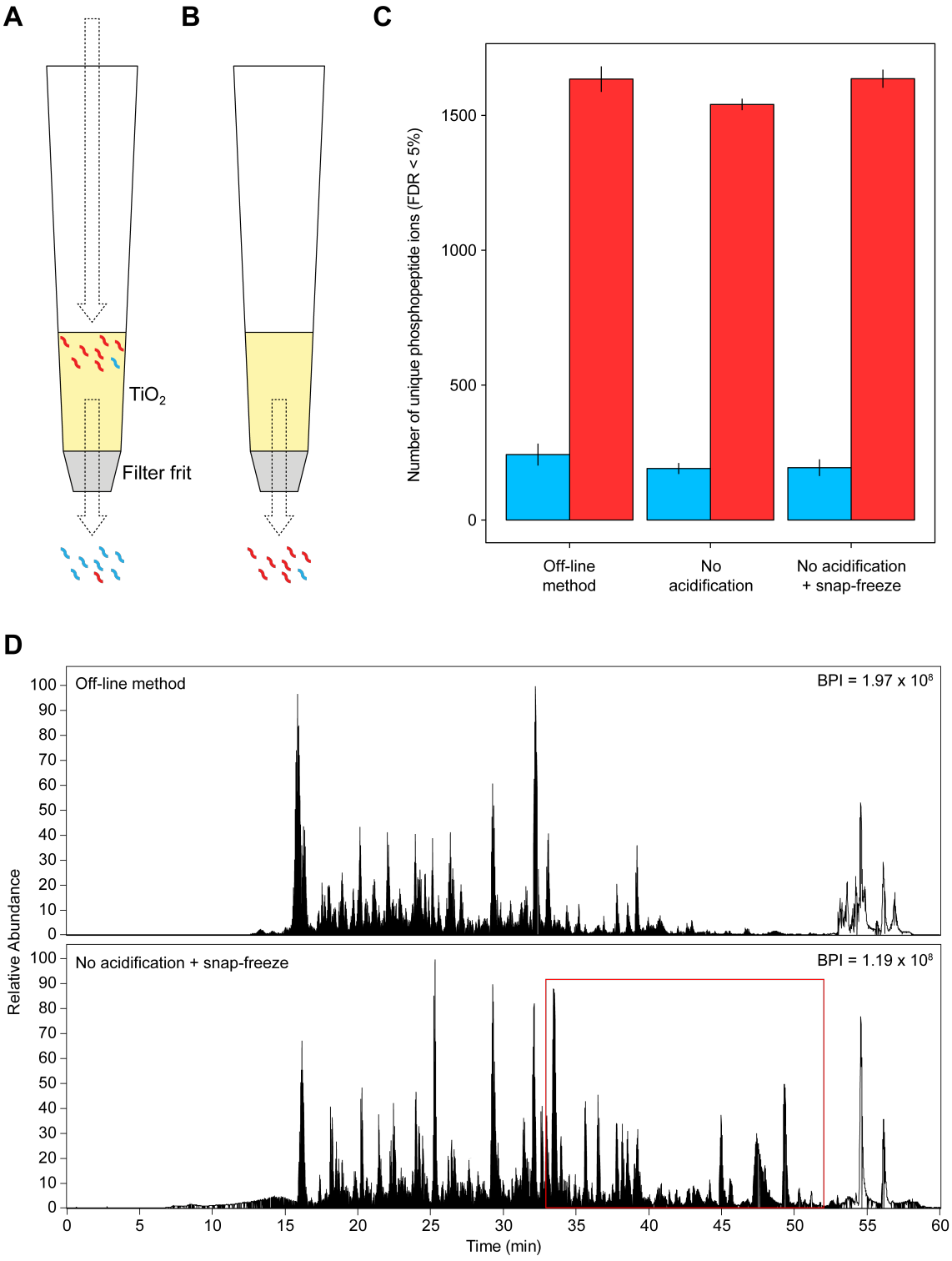


Figure 3.6: Removal of phosphopeptide acidification has no impact of phosphopeptide recovery, therefore streamlining the method. (A) Following incubation of the total peptide mixture with TiO_2, the beads are loaded into empty micro-spin columns and washed with phosphopeptide excluders. (B) The remaining phosphopeptides are then sequentially eluted with NH_4OH and immediately vacuum-dried (with or without prior snap-freezing on $CO_{2(s)}$). (C) Comparison of the number of phosphopeptides recovered via the off-line desalt method and methods involving the direct vacuum-drying of the unacidified phosphopeptide eluate $\pm$ snap-freezing. Red, phosphopeptides; blue, non-phosphopeptides. Error bars represent SEM (n = 3). *P*-values were determined via a two-tailed Student's *t*-test. (D) Representative chromatograms (normalised to the base peak) of a sample enriched using the off-line method (upper panel) and one enriched without acidification, with snap-freezing (lower panel). BPI, base peak intensity.

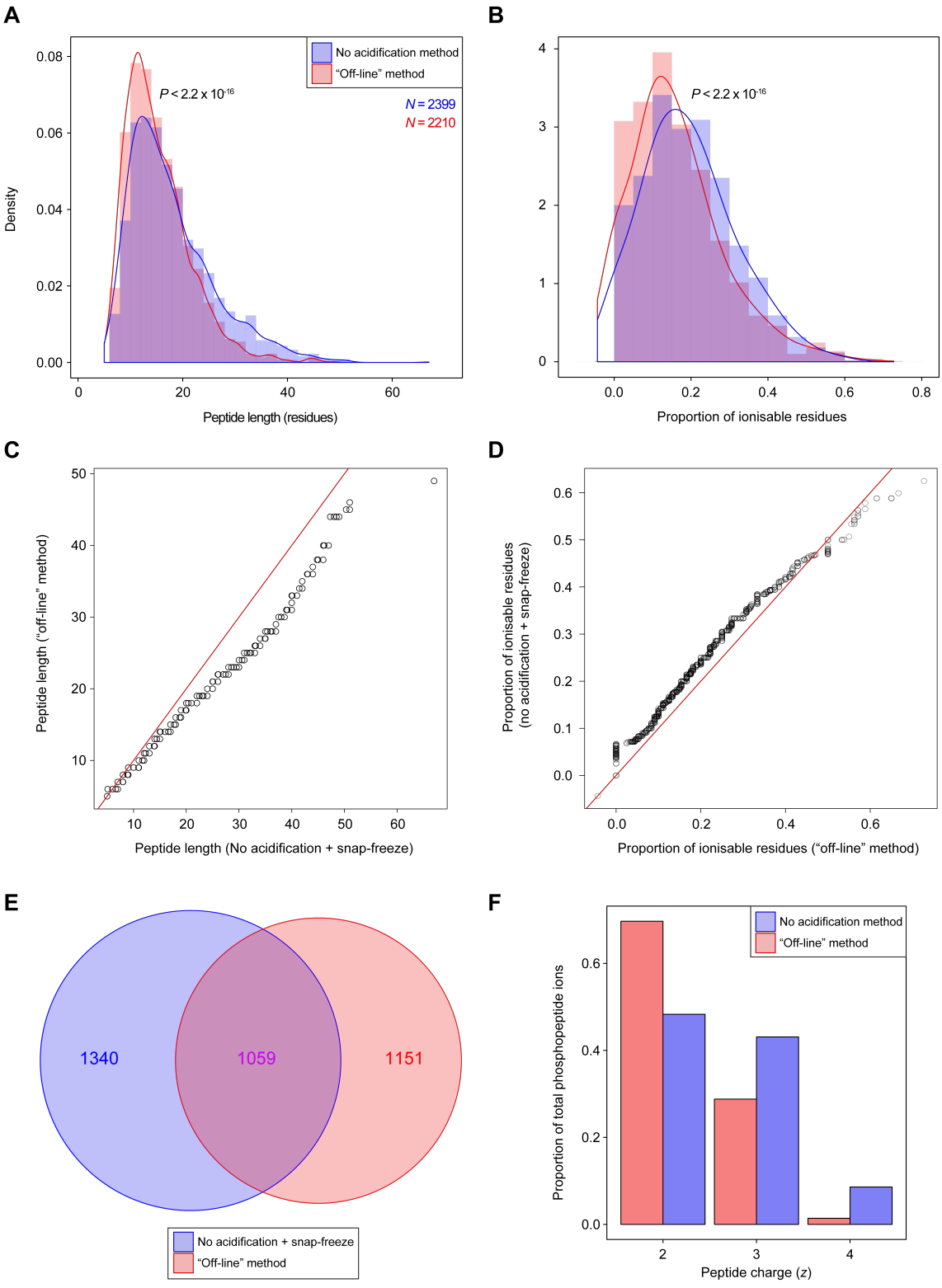


Figure 3.7: Removal of phosphopeptide acidification results in the recovery of longer and more highly charged phosphopeptides. (A) Probability density function of peptide length for combined databases of the samples processed using the off-line method (red) or those processed using the no acidification and snap-freeze method (blue). (B) Probability density function of the number of ionisable residues within each phosphopeptide for the samples as in (A). (C) Quantile-quantile plot for the peptide length distributions as in (A). Red line(s) represents $y = x$. (D) Quantile-quantile plot for the proportion of ionisable residue distributions as in (B) *P*-values were determined via a two-sample Mann-Whitney test. (E) Venn diagram showing the overlap in phosphopeptide identifications between the off-line and no acidification methods. (F) Distribution of charge states for the identified phosphopeptides, expressed relative to the total number of identifications.

steps/desalting resulted in the recovered phosphopeptides possessing higher charge states (Fig 3.7F). Taken together, these data suggest that the removal of acidification and desalting, whilst drastically streamlining the protocol (reducing potential sources of error/peptide loss) and reducing the material cost per enrichment, has no effect on the net recovery of phosphopeptides.

3.3 Comparison of mass spectrometric methods for phosphoproteomics

Alongside the importance of sample processing methods, the optimisation of mass spectrometric methods used for phosphoproteomics experiments is also key to the optimal detection and sequencing of phosphopeptide ions. As discussed in Chapter 1, such methods include altering the fragmentation technique at the MS^2 level (e.g. CID-MSA, ECD, HCD, EThcD, etc.) or altering the data-dependent selection of precursor ions (e.g. DIA or exclusion lists). The impact of these on phosphopeptide identifications will be discussed in the following sections.

3.3.1 Effects of gas-phase fractionation on precursor ion selection

Data-dependent acquisition methods are, despite ever-faster duty cycles, notorious for undersampling. A number of techniques have been developed to circumvent this issue, e.g. DIA[163,164], exclusion and inclusion lists[272], and gas-phase fractionation[273]. Gas-phase fractionation (GPF) relies on the sampling of different *m/z* windows for MS^2-selection in separate injections of the same sample. Therefore, alongside being selected based upon their intensities, precursor ions are only accepted between pre-programmed *m/z* ranges in each run (Fig 3.8A). Theoretically then, deeper phosphoproteome coverage is obtained as different populations of precursor ions are fragmented in each injection, thus reducing redundancy in the MS^2 spectra produced[273]. This method is particularly attractive with respect to MS^1-based quantification methods such as PESCAL, as the MS^1 data produced from each injection are identical (thus forming the routine quantitative technical replicates), yet phosphoproteome depth and coverage is theoretically increased with no additional analysis time.

In order to assess the impact of GPF on the number of phosphopeptide identifications, an experiment was designed whereby a single preparation of phosphopeptides was subjected to six LC-MS/MS injections on the LTQ-Orbitrap-XL mass spectrometer – three operating with full-scan MS^2 selection and three operating with GPF (window 1: *m/z* 375-625; window 2: *m/z* 625-805; and window 3: *m/z* 805-1800). These *m/z* windows were chosen on the basis that they contained equal numbers of ions based upon previously acquired MS^2 data generated from MCF7 cells. In each case, the samples were first separated on a 45 min gradient (5-35% ACN) and the ions fragmented by CID-MSA (top 3). This experiment was then repeated to produce technical duplicates. The results of this experiment are presented in Fig 3.8.

To confirm that the mass spectrometer was indeed only selecting precursor ions for fragmentation within certain *m/z* windows, the *m/z* distributions of the phosphopeptide ions identified in each injection were analysed. These data demonstrate that the instrument was indeed effective at fractionating the precursor ions selected for fragmentation, as only precursors between the specified ranges were identified in each injection when operated in GPF mode (Fig 3.8B). On the contrary, the injections operated in full-scan mode identified phosphopeptide ions across the full *m/z* range (Fig 3.8C). Despite the ability of the instrument to effectively fractionate peptide ions selected for MS^2 in the gas-phase, significantly fewer identifications were made in those samples that were run in GPF mode compared to three identical injections operating with full-scan MS^2 selection (Fig 3.8D; $P = 0.021$). Although fewer identifications were made, it was hypothesised

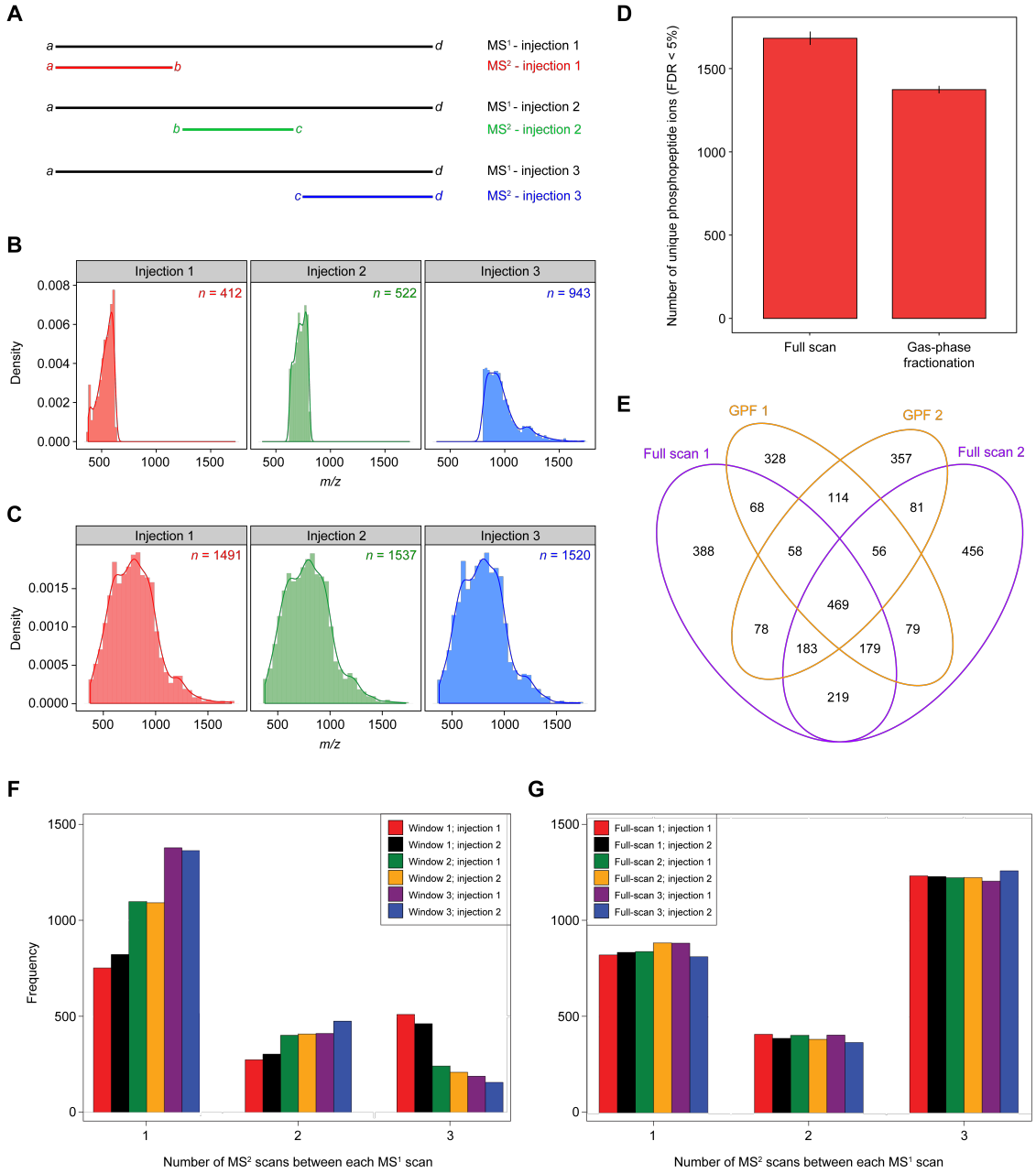


Figure 3.8: Gas-phase fractionation results in fewer phosphopeptide identifications. (A) Schematic of the principles of GPF. The sample is injected *n* times (in this case three). In each injection, identical MS^1 data are obtained (*m/z* a-d); however, different windows of precursor ion *m/z*-ratios are selected for MS^2 fragmentation in each injection (e.g. only precursor ions between *m/z* a-b are accepted for MS^2 in injection 1). (B) The distribution of identified precursor ion *m/z*-ratios in each injection when operated in GPF mode. (C) The precursor ion *m/z* distribution for each injection when operated in full-scan mode. (D) The mean number of unique phosphopeptide ions identified in combined databases of all three technical replicates for full-scan and GPF-based MS^2 precursor selection ($n = 2$). Error bars represent the maximum and minimum value. (E) Venn diagram illustrating the overlap between phosphopeptide identifications made in each set of three injections. Each circle represents a combined database of the three injections performed for each scan mode (with two replicates per scan mode). (F) and (G) Top *N* spacing distributions for the samples run with GPF and full-scan precursor selection respectively. Window 1, *m/z* 375-625; window 2, *m/z* 625-805; window 3, *m/z* 805-1800.

that in GPF mode, the instrument samples a different population of phosphopeptides, as low intensity peptides that may be ignored in full-scan mode, are selected for fragmentation. Therefore, if the instrument was indeed sampling different populations to those ions observed in full-scan mode, the overlap in identifications between the two modes would be minimal. To assess this, the overlap in phosphopeptide identifications between combined databases of the three injections for each scan mode were compared (Fig 3.8E). These data show that, as expected for technical replicates[274], the overlap in identifications between replicates in full-scan mode was 69% (Fig 3.8E, purple circles) and 75% in GPF mode (Fig 3.8E, orange circles). The overall overlap between the identifications made between GPF and full-scan mode was 74%, suggesting that the unique identifications observed in GPF mode (a total of 799 between GPF 1 and 2) were more likely to be as a result of under-sampling, as opposed to the sampling of a unique population of ions. Further analysis of the top *N* spacing distributions of samples run with GPF and full-scan precursor selection revealed that the GPF precursor ion MS^1 scans rarely triggered the maximum of three MS^2 scans, as the distribution was shifted towards only one MS^2 scan (Fig 3.8F). This was most prominent for the third *m/z* window, suggesting that this was the least populated with multiply charged peptide ions amenable for selection for fragmentation. Conversely, the majority of the full-scan injections triggered the maximum of three MS^2 scans between each MS^1 scan (Fig 3.8G).

3.3.2 Comparison of CID-MSA and HCD for phosphopeptide fragmentation

Several studies have investigated the utility of different fragmentation methods for phosphoproteomics experiments[248–250,275]. In order to optimise the fragmentation methods with those available to the Cutillas lab however, the utility of HCD and CID-MSA for single-shot, global phosphoproteomics experiments was assessed. To this end, four TiO_2 extractions (from four independent 500 μg digests derived from a pooled MCF7 lysate) were run on an LTQ-Orbitrap-Velos instrument connected on-line to a Thermo EASY-nLC. Each sample was injected four times, separated over a three hour gradient (5-35% ACN), and each injection subjected to a different ion activation method: CID-MSA (top 10), CID-MSA (top 15), HCD (top 7), and HCD (top 10). The resulting CID-MSA and HCD data were searched against the Uniprot/Swissprot human database as per previous experiments ($\pm$10 ppm precursor ion tolerance); however, the HCD data were searched with a fragment ion tolerance of $\pm$50 mmu, as opposed to $\pm$600 mmu for the CID-MSA data. The results of this comparison are presented in Fig 3.9.

The data shown in Fig 3.9A demonstrate that a significantly greater number of phosphopeptides were identified in those samples sequenced using CID-MSA compared to those sequenced using HCD, regardless of top *N* spacing (P = 0.001 [CID10 vs HCD7]; P = 0.001 [CID10 vs HCD10]; P = 0.0006 [CID15 vs HCD7]; P = 0.0005 [CID15 vs HCD10]). Interestingly, although the data show a slight trend towards an increase, the number of identifications did not increase significantly with an increase in the maximum number of MS^2 scans permitted between each MS^1 scan (Fig 3.9A). Despite the lower total number of phosphopeptides identified using HCD, a comparison of the distributions of Mascot delta scores (δ; the higher the score, the greater the confidence in phosphosite localisation) revealed that a greater proportion of the identified phosphopeptides possessed a more confident assignment of the phosphosite location – 44.1% and 53.5% possessing a $\delta > 10$ with CID-MSA and HCD respectively (Fig 3.9B). In-line with this observation, analysis of the Mascot peptide score distributions revealed that HCD produced peptides with higher average Mascot scores ($\mu_{1/2}$ = 37.82) compared to those sequenced using CID-MSA ($\mu_{1/2}$ = 30.09), therefore providing greater confidence in the assigned peptide identities (Fig 3.9C).

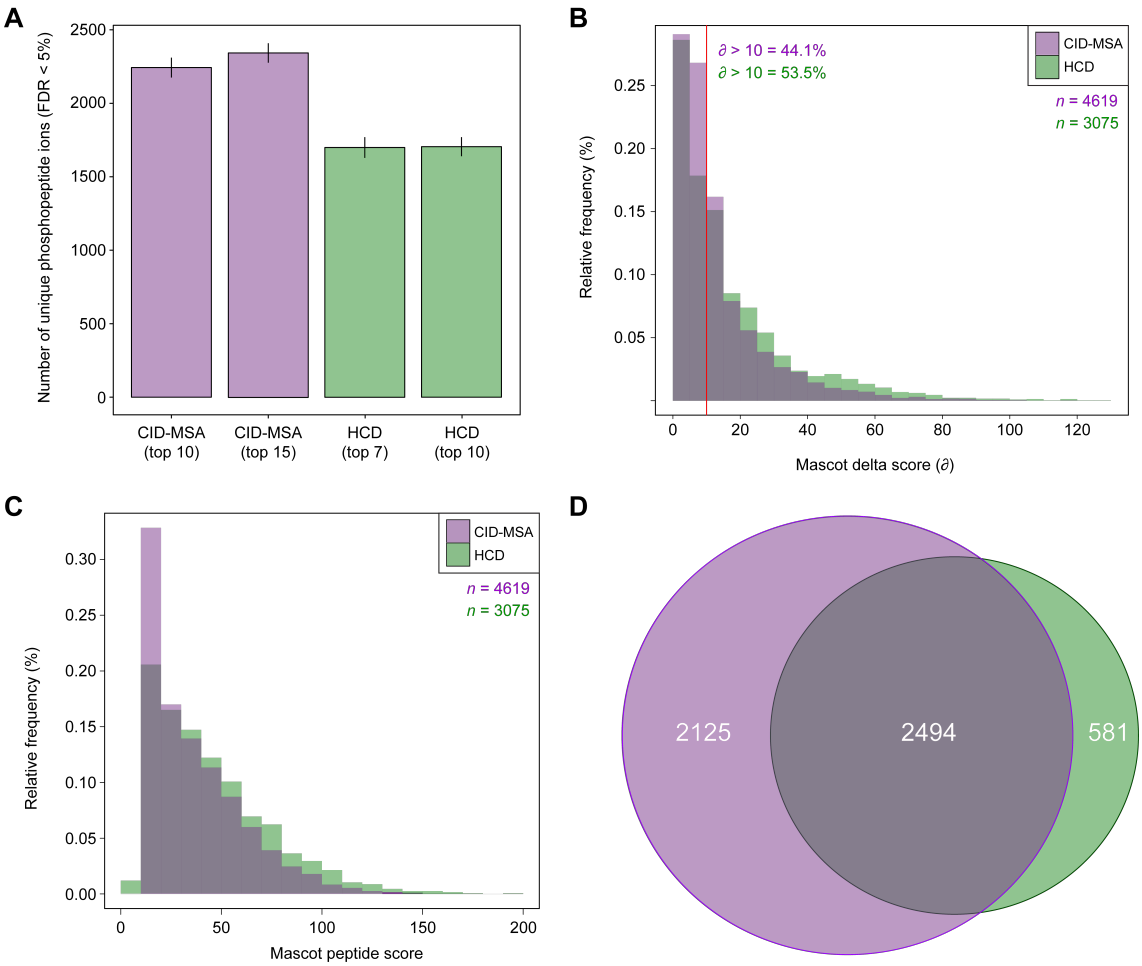


Figure 3.9: HCD ion activation identifies fewer phosphopeptides than CID-MSA. (A) The mean number of unique phosphopeptide ions identified using each ion activation method. Error bars represent SEM (n = 4). P-values were determined via a two-tailed Student's t-test. (B) and (C) Relative frequency distributions of Mascot delta and peptide scores in combined phosphopeptide databases from all the samples run in either CID-MSA or HCD mode, respectively. (D) Venn diagram illustrating the overlap in identifications made using the two fragmentation modes.

Finally, a comparison of the unique phosphopeptide ion identifications made using each method demonstrated that the large majority of those identified in HCD mode were also identified using CID-MSA (Fig 3.9D), suggesting that the differences observed were more likely to have been as a result of under-sampling in the CID-MSA runs.

3.4 Conclusions

Several modifications to the original phosphopeptide enrichment protocol were proposed and experiments designed to test their efficacy (Figs 3.1-3.7). Furthermore, several mass spectrometric methods for phosphopeptide precursor ion selection and fragmentation were compared (Figs 3.8 and 3.9). The overall aim of these investigations was to increase the depth of phosphoproteome coverage in single-shot enrichments of whole-cell lysates, increase the robustness and efficiency of the enrichment protocol, and to reduce the frequency of unwanted modifications (i.e. *N*-formylation).

Initially, a series of experiments were performed to determine the effect of an on-line desalting method on phosphopeptide recovery and subsequent modifications to this method. These data demonstrated that the on-line method resulted in the recovery a significantly larger amount of phosphopeptides and dramatically reduced the frequency of *N*-formylation (Fig 3.2). The specificity of the enrichment, however, was significantly poorer than the original method, as there was a concomitant increase in non-phosphopeptide identifications. It was subsequently hypothesised that an additional salt wash (100 mM ammonium acetate; 25% ACN) might result in an increased specificity of phosphopeptide enrichment. An experiment comparing the first iteration of the on-line method and that with an additional wash demonstrated no significant difference in the specificity of the enrichment (Fig 3.3). This, however, may have been a result of the non-phosphopeptides being removed from the TiO_2, but subsequently binding to the C18 layer due to the concentration of ACN not being sufficient to exclude this. To this end, and based on prior findings[141], the ammonium acetate wash was kept in the protocol for the later modifications.

To further investigate protocol modifications that may increase specificity whilst not reducing yield, an off-line desalting method was proposed, whereby the TiO_2 enrichment and subsequent desalting steps were decoupled. Alongside this, the effect of a more retentive sorbent (C18+GC) on phosphopeptide yield was also tested (Fig 3.4). These data demonstrated that the off-line method, combined with the C18+GC sorbent for subsequent desalting, drastically improved both the specificity of the enrichment alongside increasing the yield of phosphopeptides by approximately 1.6-fold. As a result of these data, the off-line method was taken forward for further testing.

After having demonstrated the success of the off-line method in improving both specificity and yield of single-shot phosphopeptide enrichments, a number of parameters were adjusted and their effect on the enrichment assessed. To this end, the concentration of TiO_2 in the initial incubation, the concentration of ACN in the elution buffer, and the several different TiO_2 loading buffers were optimised (Fig 3.5). These experiments showed that doubling the concentration of TiO_2 in the initial incubation, whilst not doubling the identifications that could be made, did significantly increase the number of identified phosphopeptides (Fig 3.5A). Furthermore, it was observed that a 10% increase in the proportion of ACN in the C18+GC elution buffer also significantly increased the number of identifications (Fig 3.5B). Contrary to the data published by Kettenbach and Gerber[240], the presence of TFA (5%) in the loading buffer had a positive effect on the number of recovered phosphopeptides, although an increase the concentration of the relevant α-hydroxy acid rescued this effect (Fig 3.5C). Taken together, out of the conditions tested, these data suggested that 50

μg $TiO_2 \cdot \mu$g protein^{-1}, 60% ACN, and 1M glycolic acid (80%; 5% TFA) were the optimal conditions for maximal phosphopeptide recovery.

Despite the optimisation and success of the off-line method, a more time-efficient and robust protocol with less reliance on multiple sorbents (therefore increasing the chances of peptide loss/sorbent failure and introducing potential sources of error) remained desirable. Indeed, in order to remove the reliance on the desalting step, the acidification of the phosphopeptide eluents from the TiO_2 would have to be removed (to avoid *N*-formylation). It was hypothesised, however, that due to the lack of a strong base (e.g. $Ba(OH)_2$) and lack of high temperatures experienced by the phosphopeptides at any stage of the protocol, the rate of β-elimination of phosphate groups in the NH_4OH elution buffer would be insignificant. In order to test this, an experiment was designed to compare the off-line method with a method devoid of acidification (and therefore the desalting step), with or without snap-freezing prior to vacuum drying (Fig 3.6). These data demonstrated that the number of recovered phosphopeptides was not affected (Fig 3.6C), despite there being a visibly larger phosphopeptide pellet following vacuum drying with the new method. This new method, however, resulted in the enrichment of a different subset of the phosphoproteome – this manifesting itself in the greater proportion of longer and more highly charged phosphopeptides (Fig 1.7). Due to its increased robustness, lower material cost, and more time-efficient nature, this method was taken forward for all future experiments.

Having optimised the method by which phosphopeptides can be enriched for the phosphoproteomics experiments described in this work, the efficacy of various mass spectrometric methods for the sequencing of these peptides were investigated. Firstly, the effectiveness of gas-phase fractionation was assessed. These experiments demonstrated that gas-phase fractionation, despite being applied successfully, was not as efficacious as full-scan MS^2 precursor selection for the identification of phosphopeptides (Fig 3.8). As a result, full-scan MS^2 precursor ion selection was the method employed for future work. Alongside alternative precursor ion selection modes, the methods by which precursor ions were fragmented, namely CID-MSA and HCD, were also compared. This experiment demonstrated that a greater number of identifications were made using CID-MSA; however, HCD data provided a slight increase in the proportion of identifications with high confidence phosphosite localisations (i.e. $\delta > 10$) and peptide identifications (Fig 3.9). Taken together, in order to increase the depth of single-shot analyses, CID-MSA was the fragmentation method taken forward for future experiments.

Chapter 4

Empirical determination of a kinase signalling network

4.1 Introduction and aims of the study

The results presented in Chapter 3 outlined the improvement and development of a robust and efficacious protocol for the enrichment of phosphopeptides from complex biological matrices, alongside the optimisation of mass spectrometric methods for the identification of phosphopeptides. Having finalised these methods and increased the depth of coverage of the phosphoproteome analysed in each experiment as a result, it was then possible to proceed with the empirical determination of the topology a kinase signalling network in an untargeted manner.

Kinase signalling networks, through the orchestrated flux of phosphorylation events, regulate a plethora of core biological processes (e.g. protein synthesis and responses to external stimuli)[276] and the translation of cell-cell interactions into functional outcomes (e.g. immunological and neurological signalling)[277,278]. Disruption of these networks has been shown to be implicated in the pathogenesis of several diseases. The pathways that lie within these signalling networks are often conceptualised as being linear and mono-directional. Concurrent with advances in the theoretical understanding of networks, and the technologies that allow their interrogation, however, these pathways have been better placed in the contexts in which they operate. Consequently, it has become increasingly appreciated that signalling pathways do not signal in isolation, but rather form compact, functional modules of large and highly inter-connected networks[123,279].

Two key properties of kinase signalling networks that facilitate their functional output are: their topology (i.e. how nodes are connected) and their plasticity (i.e. how the nodes can alter their interactions in response to perturbations). The highly evolved and complex topology of signalling networks confers them with the ability to change their output signals in response to changes in their environment. This phenomenon, often referred to as plasticity, has clear implications in the pathogenesis of multiple diseases (e.g. cancers and metabolic disorders), whereby cells have been shown to alter which stimuli they respond to and their signalling output (e.g. in a hypoxic environment at the core of a tumour, or in response to targeted kinase inhibition)[124].

Several methods have been developed to allow the investigation of network topology in biological systems. The majority of these involve the measurement of protein kinase/phosphatase signalling outputs and the inference of the network's topology from these data, on both small and large scales[280–282]. Mass spectrometry-based phosphoproteomics, and developments thereof, have been instrumental in increasing the tractability of such experiments, particularly label-free methods that facilitate the comparison of the large numbers of experimental conditions required.

The aim of the work presented in this Chapter was to empirically define the topology a kinase signalling network in MCF7 cells without any preconceptions of how the kinases interacted with one another. The definition of this network's topology would then allow unbiased investigation of its plasticity in these cells under different environmental conditions in an untargeted manner. To achieve this, a panel of structurally distinct, small-molecule kinase and phosphatase inhibitors was selected. The kinases targeted in the experiment were based on their known importance in cellular signalling and current clinical relevance. Critically, two inhibitors were selected for each kinase, so as to allow an estimation and reduction of the impact of the off-target effects of the compounds. Two inhibitors against the Ser/Thr phosphatase PP2A were included to provide a positive control, as the phosphopeptides reduced in abundance by the kinase inhibitors should demonstrate an increase in abundance upon treatment with inhibitors of PP2A. The global phosphoproteomes of MCF7 cells treated with these inhibitors were then quantitatively analysed by means of a label-free mass spectrometry-based phosphoproteomics strategy on an LTQ-Orbitrap-Velos instrument (Fig 4.1).

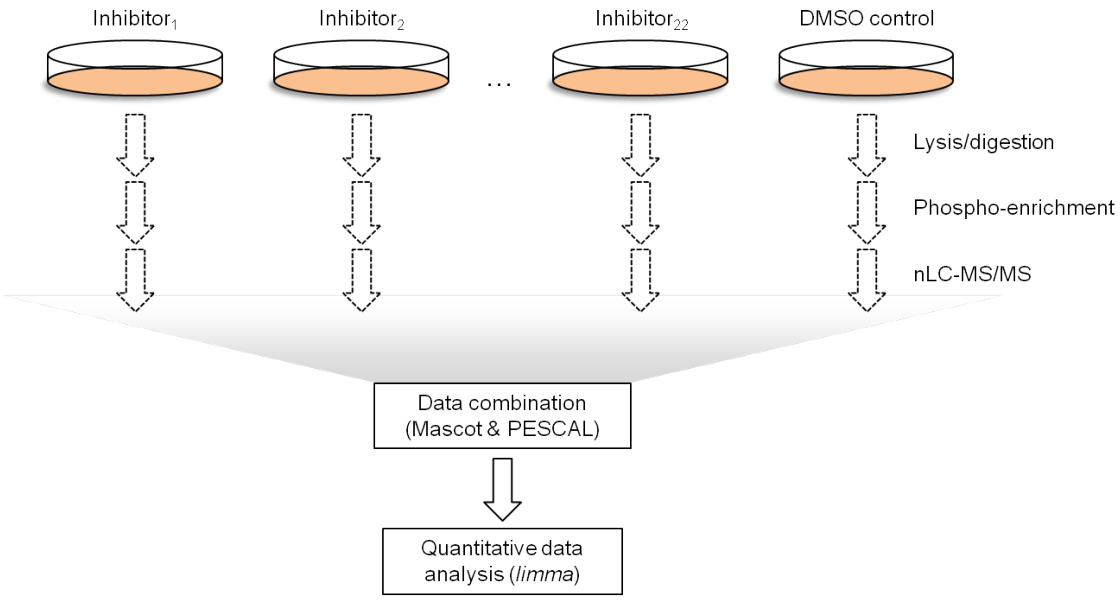


Figure 4.1: The experimental design used to empirically define the topology of a kinase signalling network. MCF7 cells were cultured and treated with one of the twenty-two inhibitors (Table 2.3) or DMSO control. The cells were then lysed, proteins digested, and phosphopeptides enriched using TiO_2 beads. Each sample was then subjected to nLC-MS/MS in triplicate. This experiment was repeated in its entirety to provide two biological replicates. The MS data were searched against the Uniprot/Swissprot database using Mascot and the resulting data combined into a single phosphopeptide database. Each phosphopeptide in the database was then quantified in each sample using PESCAL and the quantitative data analysed using *limma* (within the R statistical computing environment).

4.2 Qualitative data analysis and quality controls

Before proceeding to the quantitative analysis of the dataset produced from the experiment outlined in Fig 4.1, qualitative and quality control analyses were performed. The results of these are shown in Fig 4.2.

Combining the identifications made from each of the 144 nLC-MS/MS runs performed for this experiment produced a database of 13,405 unique phosphopeptide ions (FDR $<$ 5%). Fig 4.2A shows the distribution of *q*-values for these 13,405 phosphopeptides – this demonstrating that 95% of the phosphopeptides possessed a *q*-value $<$ 0.01, providing high confidence in their assigned identity. Out of these 13,405 identifications, 11,129 and 8,397 possessed $\delta > 5$ and $>$ 10, respectively, thus providing high confidence in their assigned phosphosite localisation.

Prior to label-free quantification using PESCAL, the sample specific retention times of each phosphopeptide ion present within the database were predicted through the use of linear models formed by pairwise alignment of spiked enolase peptide internal standard retention times. Fig 4.2B demonstrates the strongly linear relationship observed between these peptides' retention times between two independent samples. This highlighted that a linear model adequately described the relationship between two samples' enolase peptide retention times and that these peptides maintain the same relative elution order despite small retention time shifts. Assessing the strength of linear alignments between each pair of samples by means of a correlation matrix showed that this was a robust phenomenon (Fig 4.2C).

Following retention time prediction and subsequent label-free quantification of each phosphopeptide ion present within this database, the data were $\log_2$-transformed (Fig 4.2D, upper panel) and quantile normalised (Fig 4.2D, lower panel). This analysis demonstrated that the phosphopeptide intensities were approximately normally distributed (in log-space), thus lending themselves to the subsequent parametric statistical analyses. Moreover, these visualisations showed that quantile normalisation adequately normalised the intensity distributions between nLC-MS/MS

runs and conditions, thereby aiding in the removal of the majority of technical variability[265]. Finally, analysing the variability of enolase peptide retention times over the course of the 144 nLC-MS/MS runs (Fig 4.2E) revealed that, even following many sequential injections, the peptides maintained very similar retention times and identical relative elution orders. This served to provide significant confidence in the reliability of the quantitative data, as label-free methods are dependent on robust LC separations.

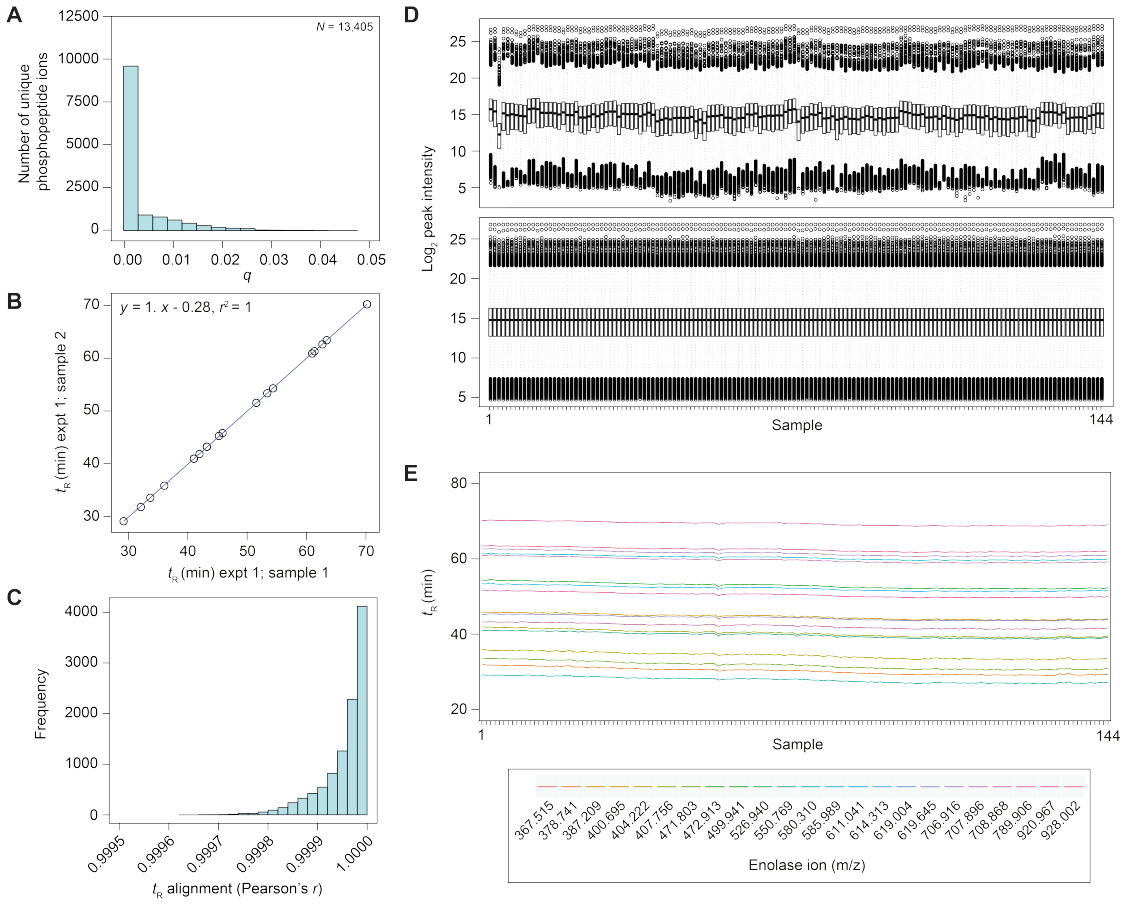


Figure 4.2: Quality control analyses of the network determination MS data. (A) Distribution of *q*-values for the identified phosphopeptides contained within the combined database created from the 144 MS runs. (B) Representative pairwise alignment of enolase ion retention times used to predict phosphopeptide retention times in samples where the phosphopeptide was not identified. (C) Distribution of Pearson's *r* values for each pairwise enolase ion retention alignment. (D) Upper panel: box plots demonstrating the distribution of $\log_2$-transformed phosphopeptide peak intensities pre-normalisation. Lower panel: $\log_2$ phosphopeptide peak intensities distributions post-quantile normalisation. (E) Enolase ion retention time vs sample (in chronological order of nLC-MS/MS run). Each line represents an individual enolase ion.

4.3 Quantitative data analysis

4.3.1 Differential abundance

Following the $\log_2$-transformation and quantile normalisation of the phosphopeptide peak intensities, the magnitude and statistical significance of changes in individual phosphopeptides' abundances were assessed using a linear modelling strategy and empirical Baye's shrinkage of standard deviations[262,266]. Multiple designs for the linear model applied to the data were tested; however, the final design (selected based on Akaike's Information Criterion) included a single DMSO control factor, a factor for each inhibitor treatment, and a factor for the biological replicates present in the experiment. The correlations between the technical triplicates were also included in

the model. Table 4.1 shows the results of this analysis, displaying the number of phosphopeptides found to be significantly up-regulated ($\log_2$ fold-ratio > 0; adjusted $P \leq 0.05$) and down-regulated ($\log_2$ fold-ratio < 0; adjusted $P \leq 0.05$). These data highlighted the large discrepancy between the proportion of the measured phosphoproteome affected by each compound, particularly regarding inhibitors targeting the same kinase (e.g. Akt inhibitor VIII and MK-2206).

Compound/inhibitor	**Reference**	**FR $<$ 0; adjusted *P* $\leq$ 0.05**	**FR $>$ 0; adjusted *P* $\leq$ 0.05**
Akt VIII	Akt^1	1784	245
MK-2206	Akt^2	684	228
KN-93	CAMK2^1	348	101
KN-62	CAMK2^2	211	152
PD-168393	EGFR^1	370	36
PD-153035	EGFR^2	162	30
ERK I	ERK^1	319	73
ERK II	ERK^2	263	37
GSK-1120212	MEK^1	68	12
U0126	MEK^2	33	15
KU-0063794	mTOR^1	295	58
Torin-1	mTOR^2	339	117
PF-4708671	p70S6K^1	373	109
DG2	p70S6K^2	343	163
GDC-0941	PI3K^1	304	137
PI-103	PI3K^2	404	175
Gö-6796	PKC^1	249	65
BIM-1	PKC^2	219	52
Okadaic acid	PP2A^1	2447	1477
Cantharidic acid	PP2A^2	457	247
H-1152	ROCK^1	267	27
Y-27632	ROCK^2	111	28

Table 4.1: The results of differential abundance analysis using *limma*. The number of phosphopeptides either significantly increased or decreased in abundance compared to DMSO are shown (n_{total} = 10,596). FR, $\log_2$ fold-ratio vs DMSO control; adjusted *P*, Benjamini-Hochberg multiple testing corrected *P*-value derived from a moderated *t*-test.

To confirm that the inhibitors were having the expected effects on the kinases they targeted, Western blots were performed for phosphorylation sites on known substrates of the target kinases/phosphatase (where reliable reagents were available). MCF7 cells were first starved for 24h in serum-free medium. The cells were then either stimulated for 1 hour with FBS, with or without the compound of interest, or left in starvation medium. The resulting data, shown in Fig 4.3, demonstrated that the inhibitors of the indicated kinases/phosphatase were having the desired effect on their downstream substrates. For example, the two mTOR inhibitors abrogated the phosphorylation of p70S6K (Thr^{389}) and the two PI3K inhibitors prevented the phosphorylation of Akt (Ser^{473}). Consistent with the greater effect on the phosphoproteome observed in quantitative MS data (Table 4.1), PP2A^1 (okadaic acid) had a more substantial effect on the phosphorylation of Akt (Ser^{473}) than PP2A^2 (cantharidic acid). These data served to highlight that these inhibitors were indeed inhibiting their target kinases in the concentrations used in the experiment, therefore providing further confidence in validity of the MS data.

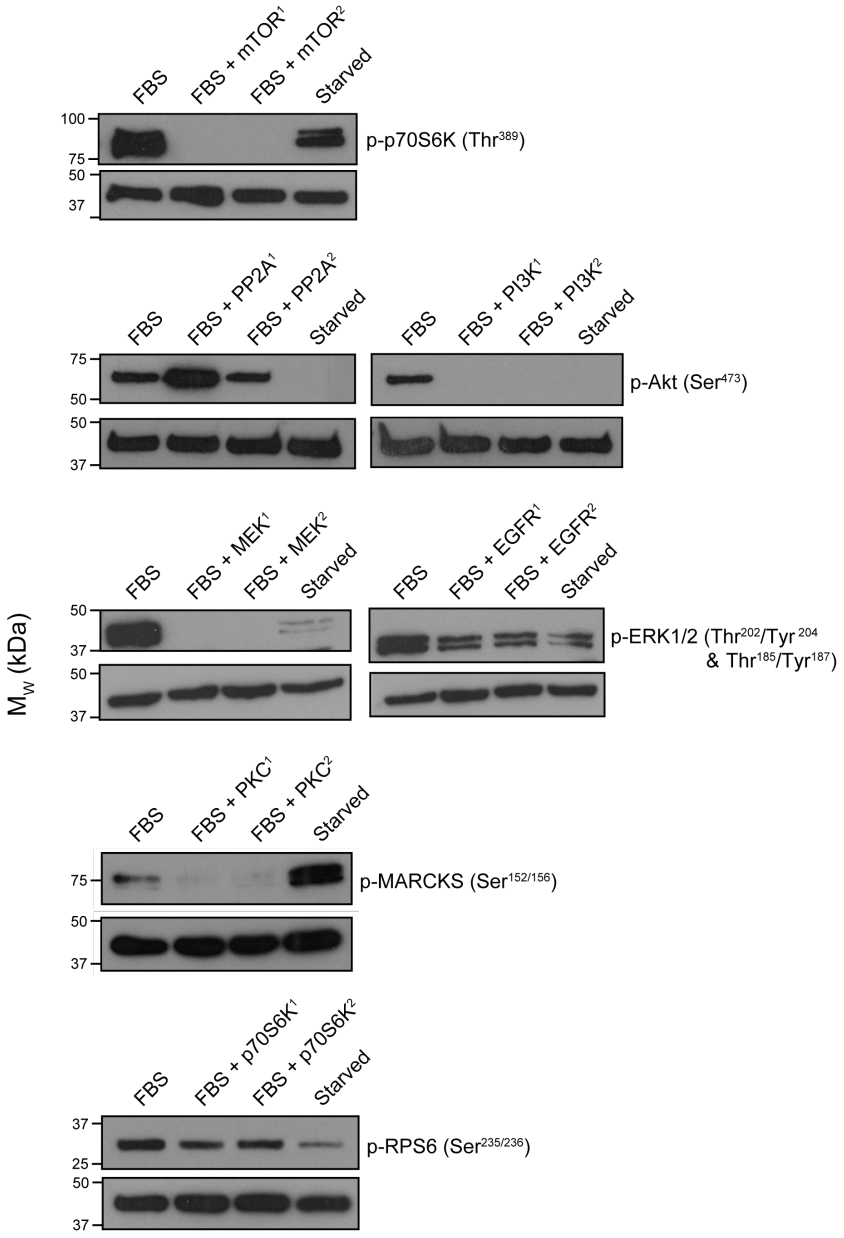


Figure 4.3: Inhibition of the target kinases/phosphatase reduces the abundance of their known downstream targets. Western blots for known downstream substrates of the indicated kinases. The loading control is total glyceraldehyde-3-phosphate dehydrogenase (GADPH) in each case. The inhibitor reference names are as shown in Table 4.1.

4.3.2 Examination of kinase inhibitor relationships

In order to determine the relationships between the kinase and phosphatase inhibitors with respects to their inhibitory effects on the MCF7 phosphoproteome, the dataset was filtered to only include phosphopeptides that demonstrated a significant (adjusted $P \leq 0.05$) decrease in abundance upon treatment with at least one inhibitor. This process resulted in a refined dataset of 4,651 phosphopeptides (out of the total 10,596 with reliable quantification data).

To analyse the relationships between the effects of each inhibitor treatment on the phosphoproteome, principal component (PC), correlation matrix, and hierarchical clustering analyses were performed (Fig 4.4). The principal component analysis of the $\log_2$ fold-ratios for the filtered 4,651 phosphopeptides revealed that, as expected, the PP2A inhibitors separated well from the kinase inhibitors in three-dimensional PC-space (Fig 4.4A). Moreover, this showed that inhibitors directed against the same kinase were closer to each other in PC-space than to the rest of the inhibitors – the only exception being the ERK inhibitors. Whilst kinase inhibitors related to the MAPK cascade (ERK, MEK, and EGFR) separated from those related to the PI3K/mTOR signalling axis (p70S6K, mTOR, and PI3K), inhibitors of Akt associated more closely in PC-space to ERK, MEK and EGFR, than to inhibitors against its well-known upstream activator, PI3K. The correlation matrix analysis reinforced the relationships observed between the Akt, mTOR, and PI3K inhibitors and between ERK, MEK, and EGFR inhibitor pairs in the PCA (Fig 4.4B; lower triangle, white boxes), as these alignments demonstrated a robust and statistically significant correlation (Fig 4.4C).

Further analysis of the mean $\log_2$ fold-ratios of the consensus phosphorylation motifs represented in the 4,651 phosphopeptides (using kinase-substrate enrichment analysis [KSEA][97]) and hierarchical clustering of the resulting data also revealed the patterns observed using PCA and correlation matrices. As such, inhibitors against the same kinase clustered more closely to one another than to the other inhibitors, and kinases of similar canonical function (e.g. p70S6K, PI3K and mTOR) clustered together. As with the other analyses, ERK inhibitors were the only exception to this observation and PP2A inhibitors clustered away from the kinase inhibitors (Fig 4.4D). Interestingly, this analysis, as observed in Fig 4.4A, demonstrated that Akt had a more similar impact on these significantly changing phosphopeptides to MAPK-associated kinases than to its canonical signalling partners mTOR and PI3K.

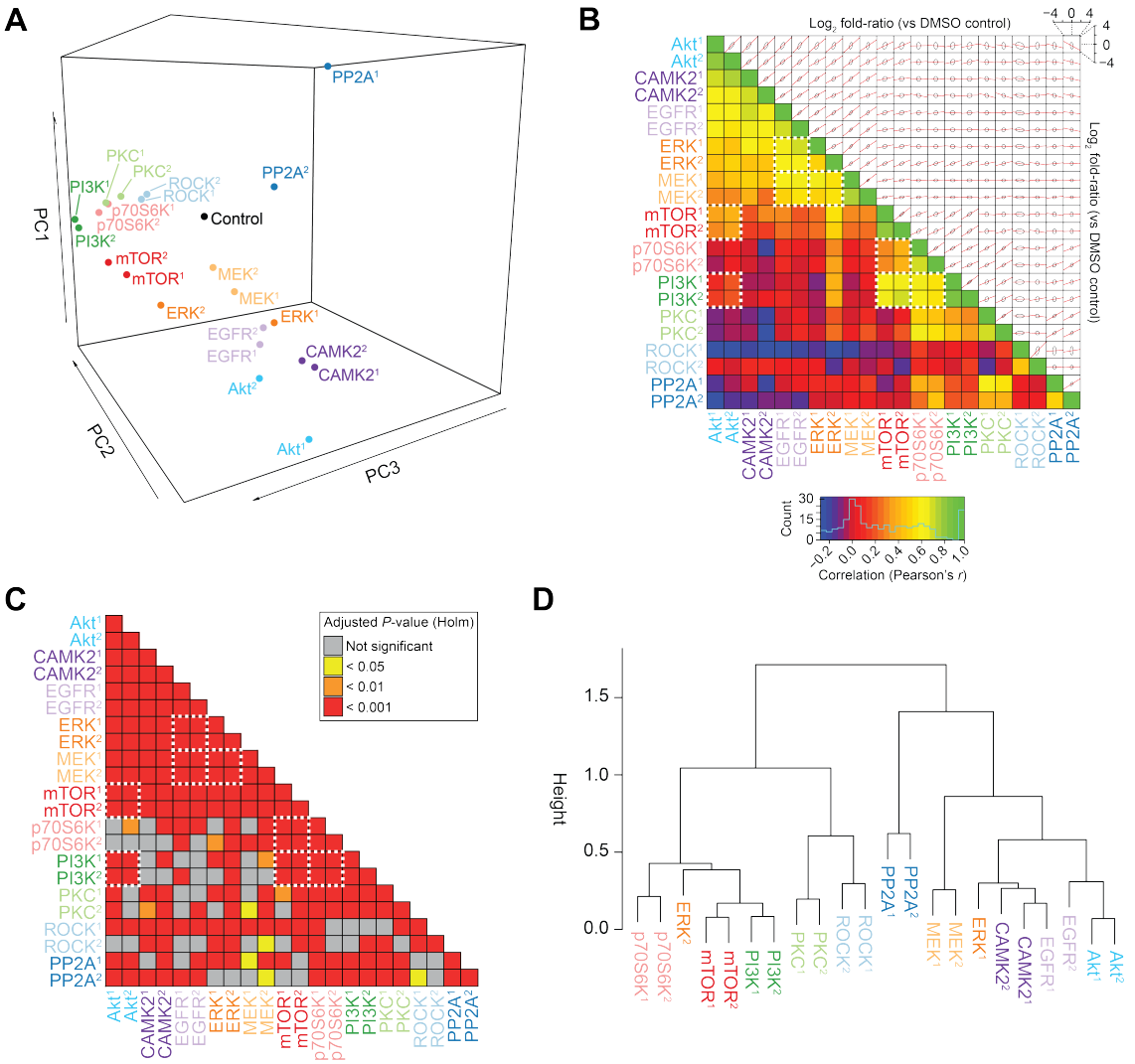


Figure 4.4: Analysis of the relationships between the effects of kinase and phosphatase inhibitors on the MCF7 phosphoproteome. (A) Principal component analysis (PCA) of the 4,651 phosphorylation sites whose abundance was significantly reduced by at least one compound/inhibitor treatment. PC, principal component. (B) Lower triangle: heatmap representing the Pearson correlation coefficients (r) between each pairwise alignment of $\log_2$ fold-ratios for each inhibitor treatment. Well-known kinase-kinase relationships are highlighted with white, dashed squares. Upper triangle: scatter plots showing the pairwise alignments between each inhibitor treatment. Red lines indicate the linear model formed between each treatment (from which the Pearson's coefficients were derived, as in the lower triangle). Grey ellipses represent one standard deviation from the mean in both dimensions. (C) Heatmap showing the significance (Holm-adjusted P-values) of each of the correlation coefficients shown in (B). (D) Hierarchical, unsupervised clustering (Pearson correlation distance metric) of the mean $\log_2$ fold-ratios for phosphopeptides containing common consensus phosphorylation motifs represented in the filtered 4,651 phosphorylation sites.

4.4 Derivation of compound-target activity markers (CTAMs) from phosphoproteomics data

In an effort to to provide an additional level of classification to the dataset, the phosphopeptides were further grouped according to their quantitative behaviour under treatment with inhibitors targeting the same kinase. To this end, it was hypothesised that phosphopeptides whose abundance was significantly reduced under treatment with both inhibitors against the same kinase (e.g. Akt[1] and Akt[2]), would represent a subset of phosphorylation sites – enriched for on-target effects of the inhibitors – which could be conceptualised as markers of both the kinase's and the compounds' activities. The identified phosphorylation sites, although not necessarily phosphorylated directly by the targeted kinase (as they could potentially be substrates of closely related kinase or those acting downstream) would be readouts of the actual kinases affected by the compound/inhibitor. These markers were therefore termed compound-target activity markers (CTAMs). Several CTAM identification criteria were tested, and the number of identified CTAMs for a least one kinase was affected as a function of these (Fig 4.5).

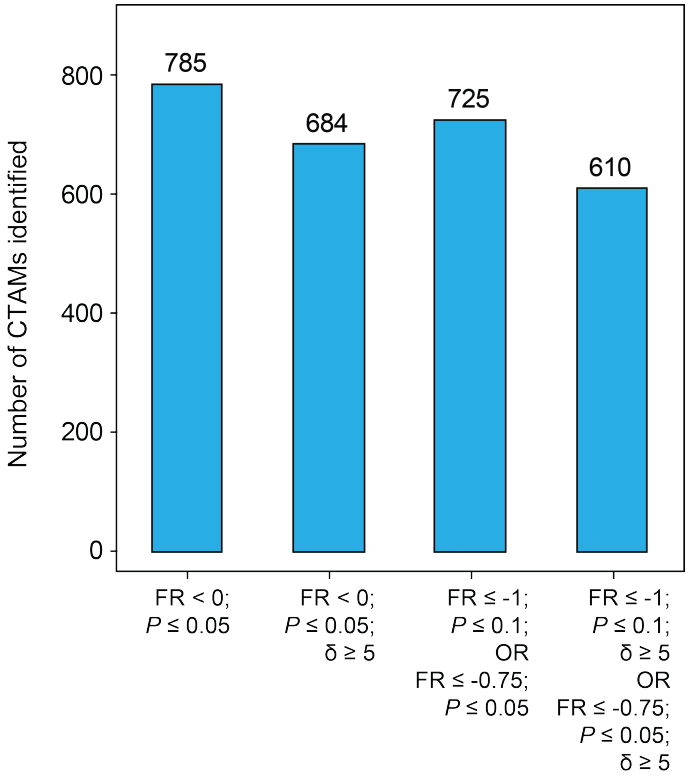


Figure 4.5: Increasingly stringent criteria reduce the number of CTAMs identified. The number CTAMs identified under each set of selection criteria. In each case the phosphopeptide in question was required to pass the thresholds in both inhibitor treatments against the same kinase. FR, $\log_2$ fold-ratio; *P*, Benjamini-Hochberg adjusted *P*-value; δ, Mascot Delta score.

As Fig 4.5 demonstrates, the number of phosphopeptides identified as being CTAMs for at least one kinase inhibitor pair decreased as the criteria were rendered more stringent (both in terms of magnitude change and statistical significance). The final thresholds were arrived at through a number of logical deductions. First, It was reasoned that, in order to reduce the number of false positive CTAMs (in terms of phosphorylation site localisation) it was deemed reasonable, despite the resulting loss of identified CTAMs, to impose a δ-score threshold. A threshold of $\delta \geq 5$ reduced the number of identified CTAMs by approximately 13-15%, but increased the confidence in the phosphorylation site localisation in the remaining peptides. Second, it was postulated that true CTAMs of the kinases targeted in the experiment would represent those that underwent sub-

stantial and robust decreases in their abundance. Therefore, phosphopeptides that possessed a $\log_2$ fold-ratio $\leq$ -1 and an adjusted $P \leq 0.1$, alongside those with a smaller ($\log_2$ fold-ratio $\leq$ -0.75) but more robust change (adjusted $P \leq 0.05$), were selected. The introduction of these extra fold-ratio thresholds reduced the number of identified CTAMs by approximately 8-10%, but increased the likelihood that the remaining CTAMs truly represented the targeted kinases'/compounds' activities. A schematic and a graphical representation of the final algorithm used to select CTAMs based on these criteria are shown in Fig 4.6A and B respectively. The data shown in Fig 4.6B demonstrate the strength of correlation between the $\log_2$ fold-ratios for the two Akt inhibitors, MK-2206 and Akt inhibitor VIII (as shown in Fig 4.4B-C). These data also illustrate, however, the number of phosphopeptides detected as being reduced in abundance by one or the other inhibitor only, suggesting that these are off-target effects of the compounds (purple boxes). In conjunction with this, these data show the different thresholds for marker selection, highlighting the necessity for the thresholds to be met for both inhibitor treatments for the phosphopeptide to be classed as a CTAM.

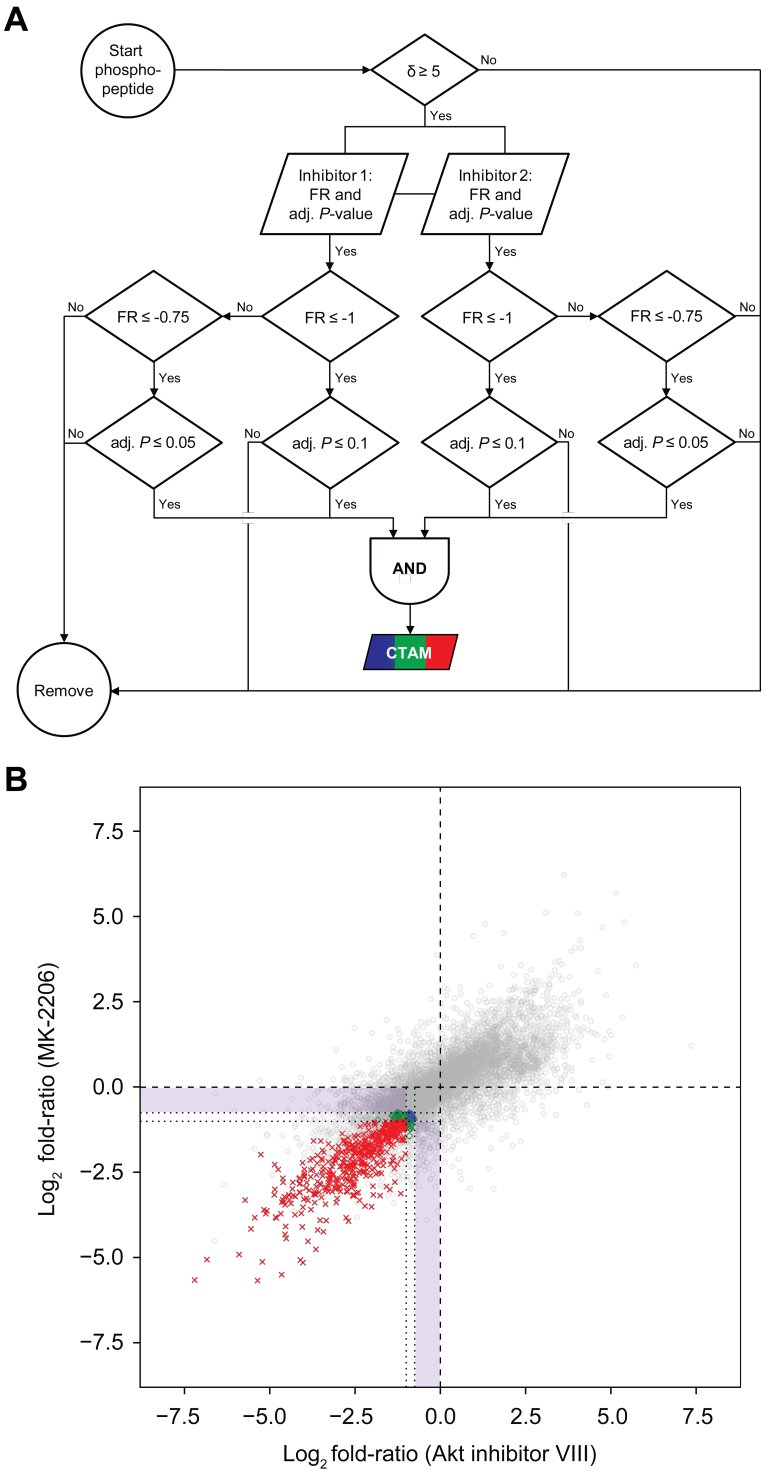


Figure 4.6: Schematic and graphical representation of the CTAM selection algorithm. (A) Schematic representation of the algorithm used for CTAM selection. Phosphopeptides (from the 13,405 identified) were serially selected. The $\log_2$ fold-ratios and adjusted *P*-values for each inhibitor pair were then used as the input. Firstly, each phosphopeptide was required to have a $\delta \geq 5$. Secondly, the $\log_2$ fold-ratios and adjusted *P*-values were assessed at several thresholds of magnitude and statistical significance. Finally, only if the phosphopeptide satisfied these criteria for both inhibitors was it defined as a CTAM. (B) Graphical representation of the algorithm described in (A). Each point represents a phosphopeptide under treatment with the two Akt inhibitors, MK-2206 and Akt inhibitor VIII. Grey, not identified as a CTAM; blue, FR $\leq$ -0.75 and adjusted $P \leq 0.05$ for both inhibitors; green, FR $\leq$ -0.75 and adjusted $P \leq 0.05$ with one inhibitor and FR $\leq$ -1 and adjusted $P \leq 0.1$ with the other; red, FR $\leq$ -1 and adjusted $P \leq 0.1$ with both inhibitors. Purple boxes represent phosphopeptides whose abundance was only reduced under treatment with one of the two inhibitors.

4.5 Inference of a kinase signalling network from phosphoproteomics data

The CTAM selection algorithm was applied to each drug pair against each kinase, resulting in the identification of 610 unique CTAMs (Appendix 1). As previously mentioned, the selection of inhibited phosphopeptides by structurally distinct compounds targeting the same kinase should have resulted in CTAM populations enriched in phosphopeptides specific to the intended kinase.

Visualisation of the 610 CTAMs simultaneously, revealed that a large number of these phosphopeptides were identified as activity markers of more than one compound-target pair. Moreover, some common patterns of CTAM identity emerged from the dataset (Fig 4.7A). This analysis demonstrated that the dataset could be utilised in this way to recapitulate known interactions between kinases (e.g. those phosphopeptides that were identified as markers of Akt, mTOR, and PI3K) alongside uncovering those that were previously uncharacterised (e.g. Akt-CAMK2). Importantly, these data could then be used to infer the topology of a kinase signalling network. To this end, the CTAMs were further classified into CTAM groups, based on the compounds/targets for which they were found to be activity markers (Appendix 2). A number of known phosphorylation site substrates of various signalling pathways emerged from this analysis and served as positive controls (Fig 4.7B). For example, 41 phosphopeptides were identified as CTAMs for Akt, mTOR, p70S6K, and PI3K. Consistent with previous knowledge, these sites included those on RICTOR (Thr^{1135}) and KS6B1 (Ser^{427})[43,61]. Similarly, sites modulated by Akt, PI3K and mTOR inhibitor pairs, independent from the p70S6K inhibitor pair, were also identified. These included sites on GSK3β (Ser^{9}), AKT1S (PRAS40 – Ser^{183}), and Myc (Ser^{62}). Furthermore, phosphorylation sites on the well-known EGFR and MEK substrates ERK1/2 (MAPK1 Thr^{185}/Tyr^{187} and MAPK3 Tyr^{204}) emerged as CTAMs of EGFR and MEK.

In addition to the recapitulation of these well-known kinase signalling cascades, evidence for the existence of uncharacterised relationships between kinases targeted by the inhibitors, and/or the inhibitors themselves were also identified. Examples of these include Akt-CAMK2 and mTOR-PI3K (independent from Akt or p70S6K) as shown Fig 4.8A-B. This analysis also provided evidence for sites that, although being identified as CTAMs of Akt, were unaffected by PI3K and other inhibitors. Alongside these, PI3K CTAMs independent of Akt and mTOR were also identified. In total, the 610 CTAMs were further grouped into 55 CTAM groups (Appendix 2).

To assess the relationships between the kinases based on their CTAMs, a PCA was performed on the $\log_2$ fold-ratios for the 610 CTAMs (Fig 4.7C). This analysis, as with those presented in Fig 4.4, showed that inhibitors against the same kinase were closer in PC-space to one another than to the other compounds and that the PP2A inhibitors (particularly the more potent $PP2A^{1}$, okadaic acid) separated clearly from the kinase inhibitors. This analysis also demonstrated that inhibitors against mTOR-PI3K related kinases and those against MAPK kinases separated in all three dimensions. Interestingly, the two Akt inhibitors were positioned more closely to the MAPK-associated kinases (e.g. EGFR and CAMK2) in the first and second PC dimensions (Fig 4.7C, upper panel); however, inhibitors of Akt were more similar to its canonical activators PI3K and mTOR in the third PC dimension (Fig 4.7C, lower panel).

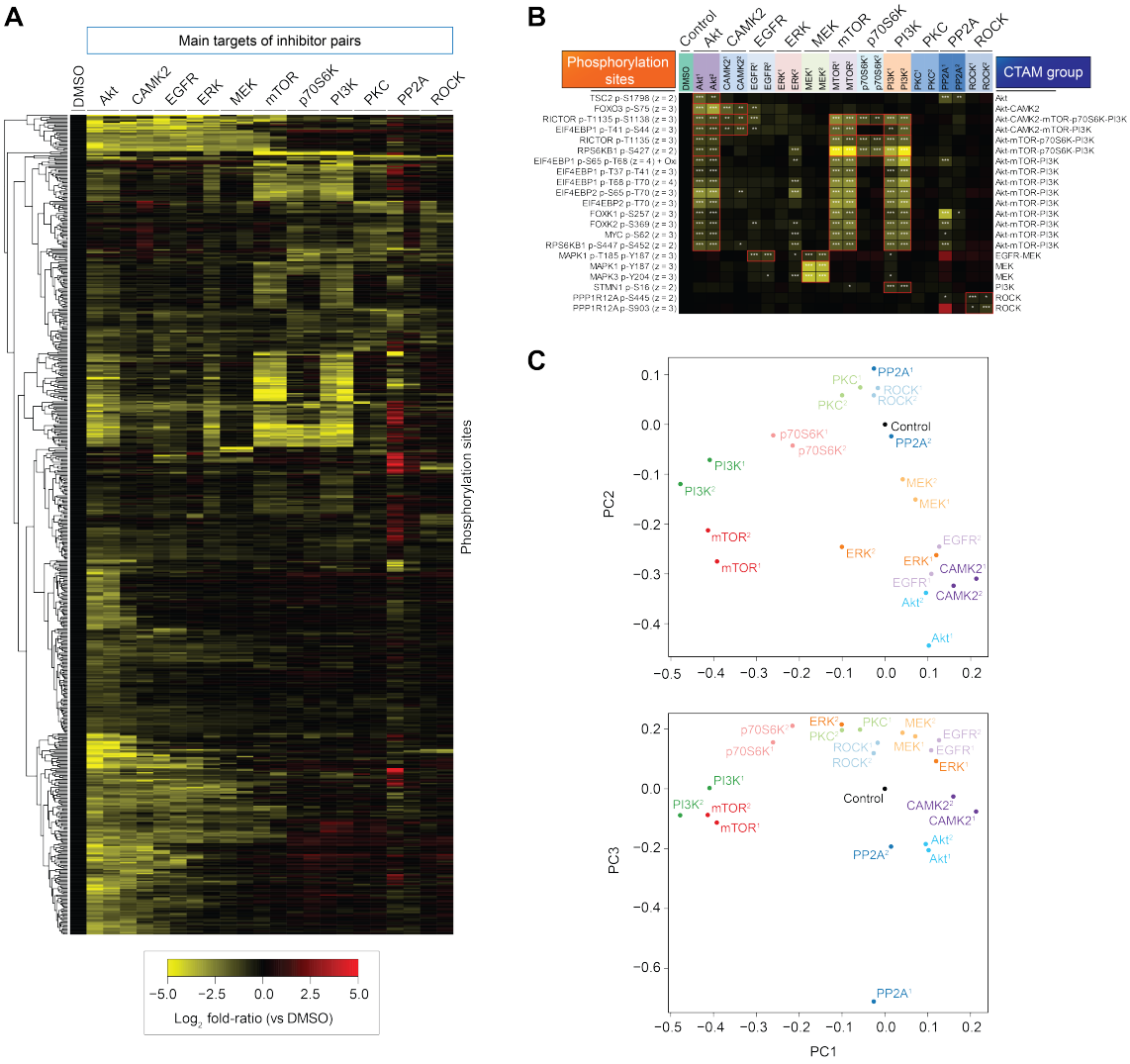


Figure 4.7: Analysis of the relationships between CTAMs and their upstream kinases. (A) Heatmap of the $\log_2$ fold-ratios for each of the 610 identified CTAMs under each inhibitor/compound treatment. The dendrogram represents the unsupervised, hierarchical clustering of the phosphopeptides according to their behaviour under each treatment (Euclidean distance metric). (B) Heatmap of selected phosphopeptides from the 610 CTAMs. Red squares highlight whether the phosphopeptide was identified as a CTAM for each specific kinase inhibitor pair. The CTAM groups each phosphopeptide was identifed as belonging to are shown (* adjusted $P \leq 0.05$, FR $\leq$ -0.75; ** adjusted $P \leq 0.1$, FR $\leq$ -1; *** adjusted $P \leq 0.01$, FR $\leq$ -1). (C) PCA of the $\log_2$ fold-ratios for the 610 CTAMs. Upper panel, PC2 vs PC1; lower panel, PC3 vs PC1.

Having defined the CTAMs and the groups of compound-target activities they represent, a kinase signalling network was then inferred from these data. To achieve this, a binary adjacency matrix was constructed. This involved the conversion of the quantitative data into a binary format, whereby an edge between each kinase node and its corresponding CTAM nodes was denoted as a “1” and those with no association as a “0”. A simplified version of such a matrix is shown in Fig 4.8A. Visualisation of this adjacency matrix as an undirected, bipartite network graph allowed a further assessment of the way in which the investigated kinases interacted with one another (Fig 4.8B).

Concordant with published data and their canonical associations, Akt, PI3K, mTOR, and p70S6K inhibitor pairs shared a large number of common CTAMs, and hence grouped together in two-dimensional space. Similarly, kinases associated with MAPK signalling (EGFR, MEK, and ERK) also grouped together in the network, alongside the related CAMK2. Most interestingly, Akt shared a large number of activity markers with both its canonical activators PI3K and mTOR, and with the MAPK-related kinases. This was observed alongside Akt possessing many activity markers that were unconnected to any of the other nodes present in the network.

A

	$Kinase_1$	$Kinase_2$	...	$Kinase_n$
$Peptide_1$	1	1	...	0
$Peptide_2$	1	0	...	1
⋮	⋮	⋮	⋱	⋮
$Peptide_n$	0	0	...	1

B

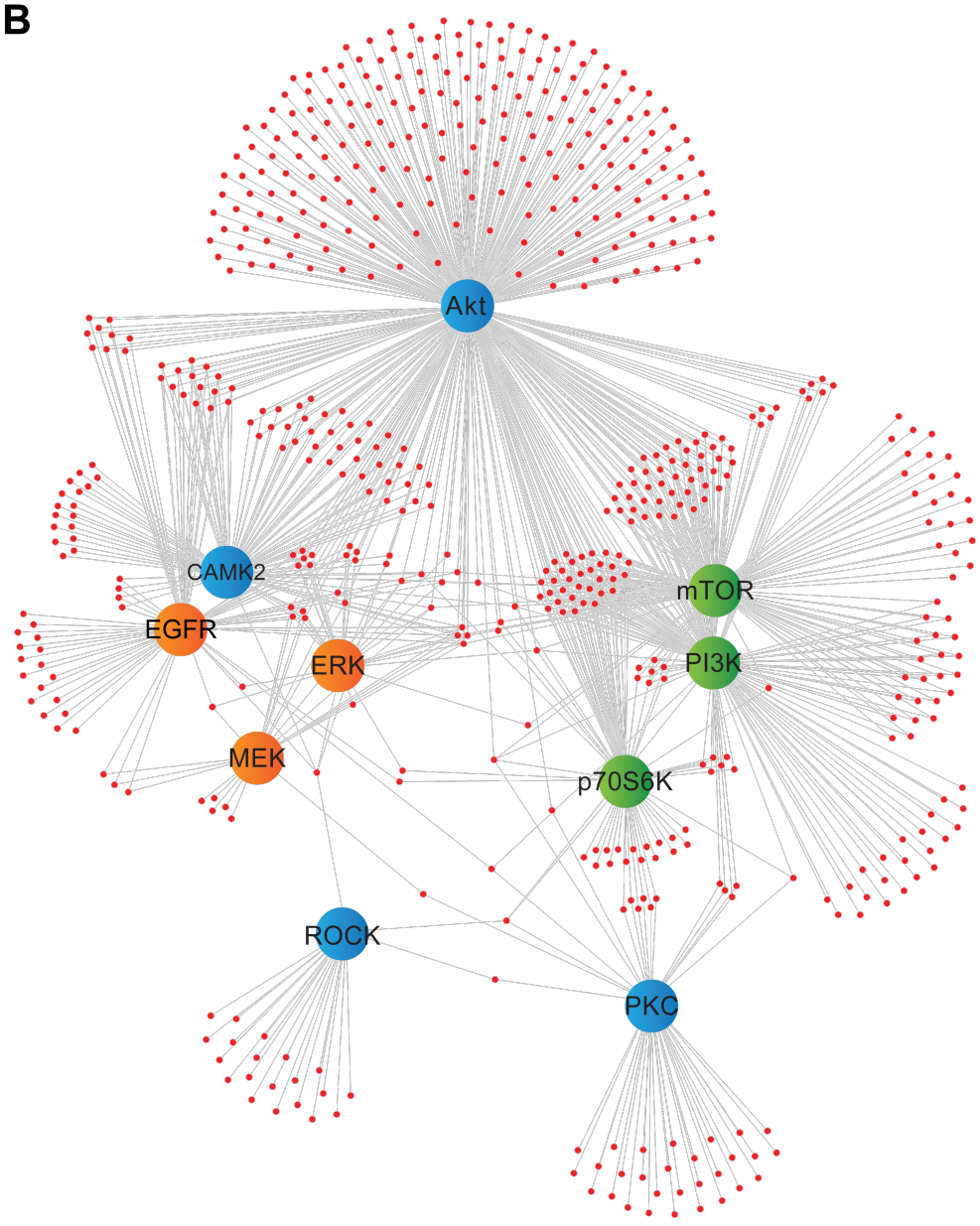


Figure 4.8: Inference of a kinase signalling network from phosphoproteomics data. Undirected, bipartite network graph representing the identified CTAMs. Large nodes represent the kinases targeted in the experiment (green, mTOR-PI3K associated; orange, MAPK-associated; blue, mixed association). Small nodes represent individual phosphopeptides, coloured according to the CTAM group they belong to. Edges between large and small nodes denote whether the phosphopeptide was identified as a CTAM for that kinase.

4.6 Conclusions

The initial objective of the experiments presented in this Chapter was to utilise the increased depth of analysis described in Chapter 3 to characterise a kinase signalling network with no prior knowledge of its topology, whilst simultaneously examining the relationships between previously well-characterised kinases in an untargeted manner. This was achieved through the use of mass spectrometry-based phosphoproteomic profiling of MCF7 cells under treatment with a panel of twenty-two structurally distinct kinase/phosphatase inhibitors (Fig 4.1).

Firstly, through a thorough quality control analysis of the dataset, it was determined that the data were of high qualitative and quantitative quality (Fig 4.2), both in terms of the identification of phosphopeptides and the subsequent prediction of their retention times and label-free quantification. Based upon previous publications advocating the efficacy of linear models for the analysis of such datasets[283,284], linear models were then fitted to the quantitative data and differential abundance of phosphopeptides (vs DMSO control) assessed by means of empirical Baye's shrinkage of standard deviations[266] (Table 4.1). This analysis showed that the magnitude in the effects of each compound on the MCF7 phosphoproteome were highly variable, as some inhibitors caused statistically significant (adjusted $P \leq 0.05$) decreases or increases in abundances of large numbers of phosphopeptides (e.g. Akt^1 and PP2A^2) and others had comparatively smaller effects (e.g. MEK^1 and MEK^2).

The effects of the inhibitors on their target enzyme was assessed by Western blot (Fig 4.3). These data showed that, of the inhibitors tested, they each demonstrated a reduction (or increase in the case of the inhibitors of PP2A) in the abundance of their known downstream substrates' phosphorylation. This suggested that the inhibitors were efficacious at inhibiting their target kinase/phosphatase, thus providing evidence to support the validity of the MS data generated using these compounds.

To further assess the relationships between the effects of each inhibitor on the phosphoproteome, the data were filtered to only include those phosphopeptides that were significantly reduced in abundance by at least one treatment. The similarity between the $\log_2$ fold-ratios of the remaining 4,651 phosphopeptides were then assessed using PC, motif, and correlation matrix analyses (Fig 4.4). These revealed that, as expected, inhibitors against the same kinase had more similar quantitative effects on the phosphoproteome than those against other kinases, suggesting that the inhibitors were indeed inhibiting the same kinases. Also, as was hypothesised, the phosphatase inhibitors had distinctly dissimilar effects from the kinase inhibitors (Fig 4.5A-D). These methods showed that inhibitors against PI3K, mTOR, and p70S6K had more similar impacts on the phosphoproteome than the MAPK-associated kinases (CAMK2, ERK, MEK, and EGFR) as these clustered separately (Fig 4.5D). Interestingly, the Akt inhibitors correlated significantly with the PI3K and mTOR inhibitors (Fig 4.5B-C); however, they clustered differently and separated in PC-space (Fig 4.5A and D).

Having analysed the relationships between different kinase inhibitors, the phosphoproteomics data were classified further in order to characterise compound-target activity markers (CTAMs) of the kinases targeted in the experiment. These CTAMs were conceptualised to represent activity markers of both the intended target kinase (both direct and indirect) and the kinases that the compounds actually inhibited. In an effort to deconvolute the off-target effects of the compounds, an algorithm was designed to select markers based upon their quantitative behaviour with respects to both inhibitors targeting the same kinase (Figs 4.5 and 4.6A). To this end, filters on the accuracy of phosphorylation site localisation, statistical significance and magnitude of inhibition were applied to reduce the presence of false positive identifications (Fig 4.5). This algorithm was assessed

graphically (Fig 4.6B) and, when applied to each kinase inhibitor pair, led to the identification of 610 CTAMs for at least one kinase/compound pair.

It was observed that a number of the identified CTAMs were identified as markers of more than one kinase/inhibitor pair (Fig 4.7A). Therefore, in order to infer signalling network circuitry from these data, these CTAMs were grouped into 55 distinct CTAM groups (e.g. phosphopeptides that were simultaneously identified as CTAMs of Akt, mTOR, and PI3K, were classified into the Akt-mTOR-PI3K CTAM group). A number of well-known pathways could be recapitulated from this analysis, such as the Akt-mTOR-PI3K and EGFR-MEK pathways (Fig 4.7B). Alongside these well-known signalling routes, a number of novel associations were also identified (e.g. Akt-CAMK2, and mTOR-PI3K bereft of Akt). PCA of the $\log_2$ fold-ratios of the 610 markers revealed similar separation of the kinases in PC-space as those observed in Fig 4.4A.

Following the identification of these markers and the groups they were classified into, the topology of the kinase signalling network they represented in MCF7 cells was inferred. The network was constructed by conversion of the CTAM membership data into a binary adjacency matrix (Fig 4.8A) and subsequent drawing of this in two-dimensional space as an undirected, force-directed, bipartite network graph (Fig 4.8B). Visualisation of the data in this manner revealed that as would be expected from their canonical association with one another, mTOR, PI3K, and p70S6K shared a large number of CTAMs. Furthermore, the MAPK-associated kinases also shared a number of CTAMs with each other and thus formed a cluster in two-dimensional space. As would be expected from canonical knowledge, Akt shared a large number of markers with its well-known upstream activators mTOR and PI3K; however, it also shared a large number with the MAPK-associated kinases and CAMK2. Moreover, Akt possessed a large number of CTAMs that were not shared with any of the other kinases targeted in the experiment, perhaps suggesting a more substantial role outside of PI3K/mTOR and MAPK signalling than has previously been appreciated.

In summary, mass spectrometry-based phosphoproteomics data was used to define a kinase network signalling topology in MCF7 cells. These experiments were able to recapitulate both known interactions between well-known kinases and signalling pathways, whilst also uncovering novel relationships between these. It remained to be proven, however, whether this network was biologically relevant and whether it would provide a meaningful output when its activity was assessed under functional activation of the kinases that it represented.

Chapter 5

Analysis of signalling network plasticity in response to network perturbation

5.1 Introduction and aims of the study

Chapter 4 outlined the derivation of compound-target activity markers, and the subsequent definition of a kinase signalling network based upon their interactions with one another, in MCF7 cells. Having finalised the topology of the network, it was then hypothesised that it could be used to assess the global signalling activity of MCF7 cells under any given treatment, without bias towards which kinases and pathways would be be activated/deactivated. Theoretically then, the cells could be treated with a stimulus, subjected to nLC-MS/MS-based phosphoproteomics and the activity of the kinases/pathways within the network determined computationally (as shown in Fig 5.1). This approach could then be used to monitor changes in the activity the defined kinase signalling network (i.e. its plasticity) upon acquisition of resistance to chronic treatment with targeted kinase inhibitors.

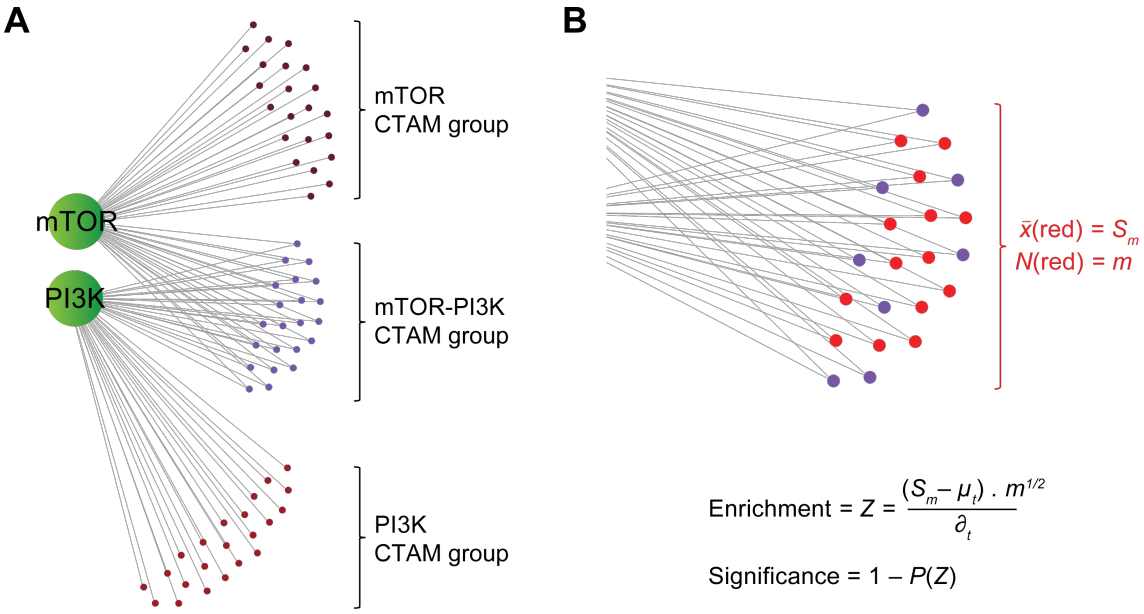


Figure 5.1: Untargeted assessment of network activity through the quantitative monitoring of CTAM groups. (A) A selected portion of the defined CTAM network as shown in Fig 4.7B. (B) In a separate experiment, MCF7 cells are subjected to an arbitrary treatment and their phosphoproteomes analysed by nLC-MS/MS. The number of CTAMs from each group identified in this experiment is then determined (m, red). The mean $\log_2$ fold-ratio (vs control) of these is then computed (S_m). The enrichment/activity of each CTAM group is calculated by means of a Z-score (μ_t = mean abundance of all phosphopeptides in the experiment; δ_t = standard deviation of all phosphopeptides). The statistical significance of this enrichment is then calculated by computing the probability that a random variable drawn from a standard normal distribution will be less than or equal to Z (as in refs[97,285])

Epidermal growth factor (EGF) and insulin-like growth factor-1 (IGF-1) are important molecules in the control of cell growth, metabolism, and proliferation. EGF binds to its cognate receptor EGFR, which, as discussed in Section 1.1.4, causes EGFR's autophosphorylation and subsequent acute activation of several downstream intracellular signalling pathways (e.g. PI3K/Akt and MAPKs)[286]. In contrast, IGF-1 binds to IGF-1R and the insulin receptor (InsR), although with significantly lower affinity for the latter. IGF-1 is one of the most potent activators of the PI3K/Akt/mTOR signalling axis and as such, serves to promote growth in a diverse array of cell-types[287]. In both cases, once bound to their ligand, the receptors undergo clathrin-mediated endocytosis, are trafficked through the cell, and are eventually degraded or recycled back to the plasma membrane[288]. Consequently, the activation kinetics of the signalling pathways downstream of these receptors follow distinct temporal patterns; MAPK signalling showing an acute increase upon stimulation and Akt signalling demonstrating a slower activation with regards to EGF, for example[286]. Due to their roles in cell growth and proliferation, both of these growth fac-

tors (and their receptors) have been implicated in the development and maintenance of cancers and other metabolic diseases[289,290].

To test whether the network was indeed biologically relevant and could be used to assess network plasticity systematically, an experiment was designed whereby MCF7 cells were starved (0% FBS DMEM) and then treated with 100 ng·mL^{-1} of either recombinant human epidermal growth factor (EGF) or insulin-like growth factor-1 (IGF-1), for 0, 5, 10, 30, or 60 minutes. The resulting lysates were then processed, phospho-enriched and analysed by nLC-MS/MS in triplicate. The experiment was repeated to produce biological duplicate measurements.

It was postulated that, if the network embodied a true representation of signalling within these cells, functional activation of the kinases within the network (with growth factors) would result in measurable and biologically meaningful pertubations within this network. Thus, the aim of the work shown in this Chapter was to determine whether the output from the CTAM network analysis recapitulated what is known about the functional activation of these kinase/pathways upon growth factor stimulation. A quality control analysis for the resulting dataset is shown in Appendix 3.

5.2 Quantitative data analysis

5.2.1 Differential abundance of phosphopeptides over the growth factor time-courses

The quantitative data resulting from the experiment were processed as outlined in Chapters 2 and 4. The final linear model design (selected based on Akaike's Information Criterion) included a single t = 0 min control factor, a factor for each time-point, and a factor for the biological replicates in the experiment. The correlations between the technical triplicates were also included in the model. Table 5.1 shows the results of this differential abundance analysis, displaying the number of statistically significant increases and decreases in phosphopeptide abundances. These data highlighted the differential effects of EGF and IGF-1 on the phosphoproteome at a global level. The number of significantly decreased phosphopeptides was much greater for those time-points under IGF-1 treatment compared to those under EGF treatment. Moreover, the difference in the kinetics of responses was evidenced in the greater number of significantly changing phosphopeptides at longer time-points under IGF-1 treatment.

Condition/time-point	**FR $<$ 0; adjusted *P* $\leq$ 0.05**	**FR $>$ 0; adjusted *P* $\leq$ 0.05**
EGF (t = 5 min)	260	741
EGF (t = 10 min)	245	862
EGF (t = 30 min)	77	423
EGF (t = 60 min)	42	348
IGF (t = 5 min)	349	792
IGF (t = 10 min)	481	779
IGF (t = 30 min)	1014	796
IGF (t = 60 min)	708	621

Table 5.1: The results of differential abundance analysis using *limma*. The number of phosphopeptides either significantly increased or decreased in abundance compared to t = 0 min are shown (n_{total} = 11,029). FR, $\log_2$ fold-ratio vs DMSO control; adjusted *P*, Benjamini-Hochberg multiple testing corrected *P*-value derived from a moderated t-test.

5.2.2 Examination of positive controls and validation of MS data

To ensure the mass spectrometry data were of good quality and to provide confidence in the downstream analyses, a number of positive control phosphorylation sites were extracted from the data and their temporal profiles analysed (Fig 5.2). As expected, the MS data showed that upon

EGF stimulation, the MAPK signalling axis underwent substantial increases in phosphorylation of its members on phosphorylation sites responsible for their activation – the upstream members RAF1 (Thr^{641}) and SOS1 (Ser^{1134}) and downstream members MEK2 (MAP2K2 Ser^{226}), ERK1/2 (MAPK1 Tyr^{187}/MAPK3 Tyr^{204}), and p90RSK (RPS6KA1 Ser^{380}), for example. Similarly, members of the PI3K/Akt/mTOR signalling axis also demonstrated the expected temporal kinetics upon both EGF and IGF-1 stimulation. For example, a phosphorylation site (most probably Ser^{263}) on the p85β PI3K regulatory subunit (PIK3R2) demonstrated an increase in abundance under both EGF and IGF-1 stimulation. Interestingly, a phosphorylation site on the closely related p150 regulatory subunit (PIK3R4), which interacts with Vps34, also increased in abundance, although this was restricted to IGF-1 stimulation. Consistent with an activation of PI3K, a phosphorylation site within the activation loop of Akt1 (residues 466-480) demonstrated a sharp, significant increase in abundance following 5 min of EGF treatment; however, consistent with IGF-1 being a more potent activator of this pathway, this increase showed more sustained kinetics with IGF-1 stimulation. In-line with these observations, the abundance of Ser^{9} on GSK3β (a direct Akt substrate) also increased under both treatments, although with different kinetics between the two. Further downstream from Akt, phosphorylation sites on mTOR substrates also increased in abundance, demonstrating an increase in mTOR activity that would be expected under growth stimulatory conditions (e.g. RICTOR, p70S6K [RPS6KB1], and eIF4-binding proteins).

To validate the activation of MAPK and PI3K/Akt/mTOR signalling observed in the MS data generated from this experiment, Western blots were performed for well-known activatory phosphorylation sites on both Akt and ERK1/2. The results of this analysis are shown in Fig 5.3. These data, as with those presented in Fig 5.2, showed that EGF and IGF-1 stimulation resulted in acute increases in the phosphorylation of Akt and ERK1/2 within 5 minutes of treatment. Moreover, in agreement with the MS data, the increase in the phosphorylation of Akt in its activation loop was sustained for a longer time period under IGF-1 stimulation. Taken together, these analyses provide evidence to support the notion that the MS data were representative of the canonical effects of both EGF and IGF-1 in these cells.

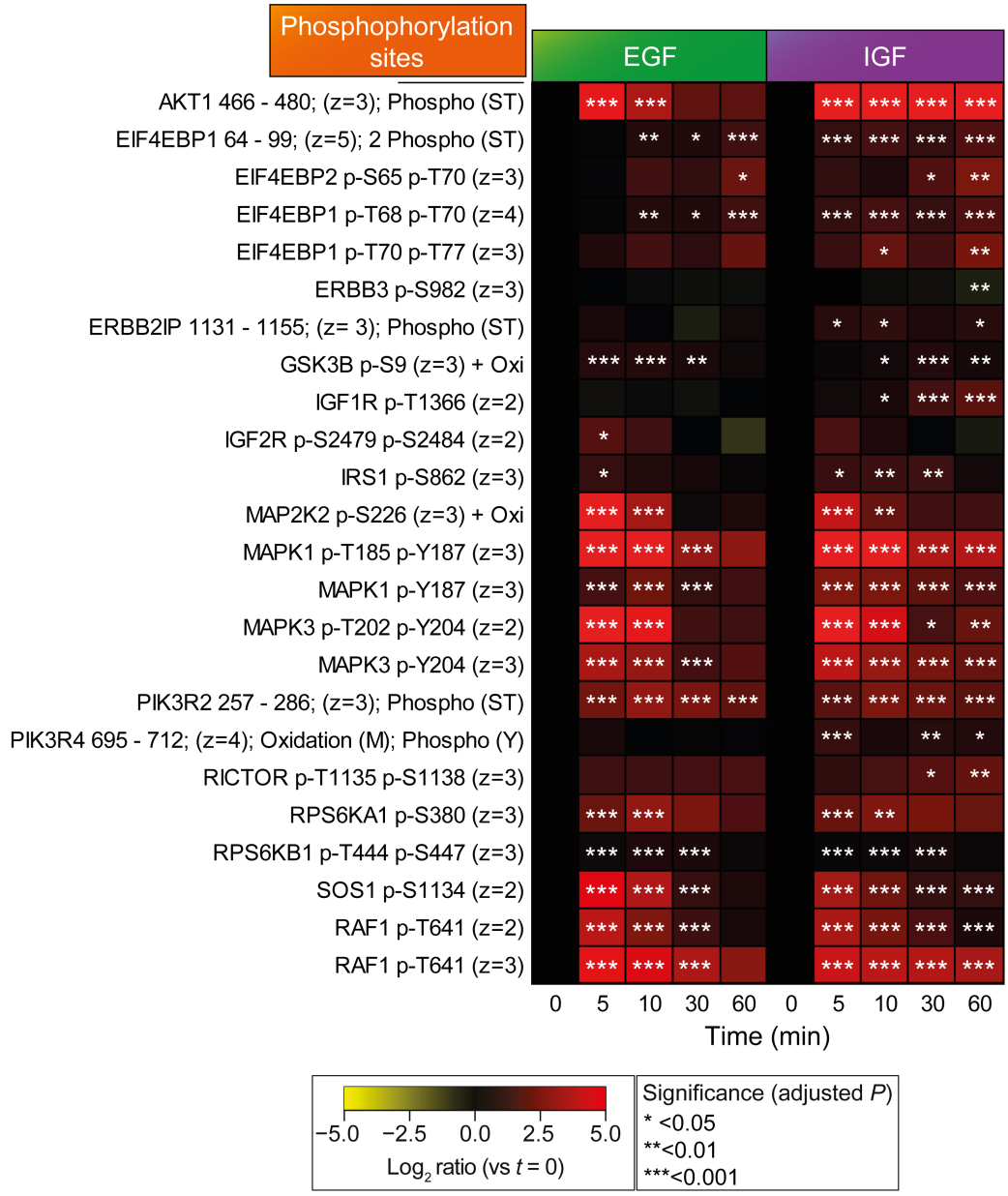


Figure 5.2: Analysis of known downstream phosphorylation sites reveals the expected effects on core signalling axes upon EGF/IGF-1 stimulation. Heatmap of the $\log_2$ fold-ratios (vs t = 0 min) for phosphorylation sites known to be downstream of the PI3K/Akt/mTOR and MAPK signalling axes. Uniprot protein ID (gene name), phosphorylation sites, peptide charge, and other peptide modifications are shown. If the δ-score for the peptide is < 10, the peptide is shown in the following format: ”protein gene name (sequence residues); charge; modification”.

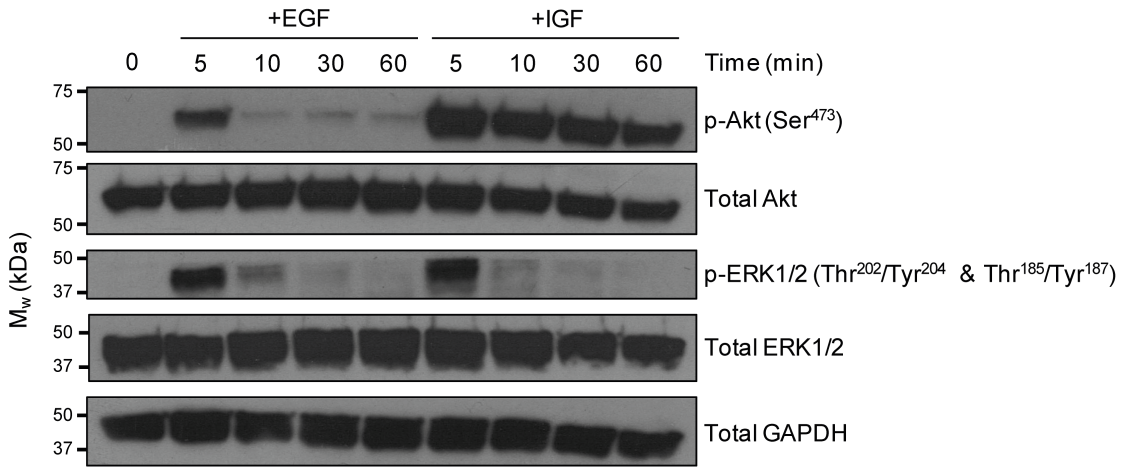


Figure 5.3: Validation of known downstream phosphorylation sites by Western blot confirms the MS data observations. Western blots for activatory phosphorylation sites on Akt and ERK1/2, as indicated. Quantity of the unmodified proteins and loading control are also shown.

5.3 Analysis of network plasticity

5.3.1 Temporal behaviour of CTAM groups

Having established that the dataset provided a veritable representation of the EGF and IGF-1-stimulated signalling in these cells, it was then sought whether the CTAM group network could be used to measure the biochemical activation of signalling pathways within the network, in an untargeted manner. It was hypothesised that, should this network provide a reliable readout of signalling activity, this analysis would show the expected signalling changes when the network was perturbed under these well-characterised experimental conditions. The CTAM network analysis was therefore performed on this dataset and the results of this are presented in Fig 5.4.

It was observed that the temporal dynamics of CTAM groups commonly associated with IGF-1R and EGFR signalling were consistent with previously reported effects of IGF-1 and EGF on kinase signalling (Fig 5.4A-B)[286]. For example, in-line with the known activation dynamics of the MAPK signalling axis and the Western blot analyses (Fig 5.2-5.3), the EGFR-MEK and MEK CTAM groups underwent acute and significant up-regulation with both EGF and IGF-1 stimulation. As Fig 5.4 shows, these groups peaked in intensity at 5 min before beginning to return to baseline after 60 min. Alongside these groups, those characterising PI3K/Akt/mTOR signalling also demonstrated the expected dynamics. The Akt-PI3K, Akt-mTOR-PI3K, and Akt-mTOR-p70S6K-PI3K CTAM groups demonstrated a steady increase in activity over the period of stimulation, peaking in signal at 60 min. Despite both growth factors having a stimulatory impact on these groups, consistent with the notion that IGF-1 is a more potent activator of these network branches than EGF, several CTAM groups demonstrated a greater intensity of activation under IGF-1 treatment (e.g. Akt-mTOR, PI3K, Akt-PI3K, and p70S6K-PI3K).

Visualisation of these data in the form of a dotplot more efficiently showed how they could be used to provide a snapshot of the network's activity under any given condition. Hierarchical clustering of the time-dependent behaviour of each CTAM group revealed which network branches signalled in a similar manner to one another in the network (Fig 5.5), and as observed in Figs 5.2-4, this demonstrated that the core MAPK signalling axis demonstrated similar behaviour between EGF and IGF-1 (EGFR-MEK and MEK groups). In addition, visualising the data in this way facilitated an easier comparison of the differences between the two growth factors; several network branches showing different temporal behaviours (e.g. Akt-CAMK2-EGFR-p70S6K and p70S6K-PI3K-PKC-ROCK). To determine more objectively whether the CTAM network model was able to resolve the differences between the two growth factors, a principal component analysis was performed on the data shown in Figs 5.4 and 5.5 (Fig 5.6).

The PCA presented in Fig 5.6 shows that the behaviour of the network as a whole was very similar between EGF and IGF-1 at 5 minutes following stimulation, and this was the time-point at which the network was most different from 0 minutes (as both points are the furthest from $t = 0$ min, yet very similar to one another in PC-space). The state of the network at 10 minutes then began to highlight the differences between IGF-1 and EGF regarding their effects on the cells' signalling (as these began to separate in the second PC dimension). The plot then demonstrates that the network began to relax towards its basal state (i.e. returning towards 0 in both dimensions); however, the way in which it proceeded to do this was different between the two growth factors. In this regard, the network state under EGF stimulation at 60 minutes appeared to be closer to 0 minutes in PC-space than at 60 minutes with IGF-1, suggesting that the activity of the network returned towards its basal state more rapidly than under treatment with the latter. This was again consistent with the prior notion, and evidence observed in the other analyses, that IGF-1 has a

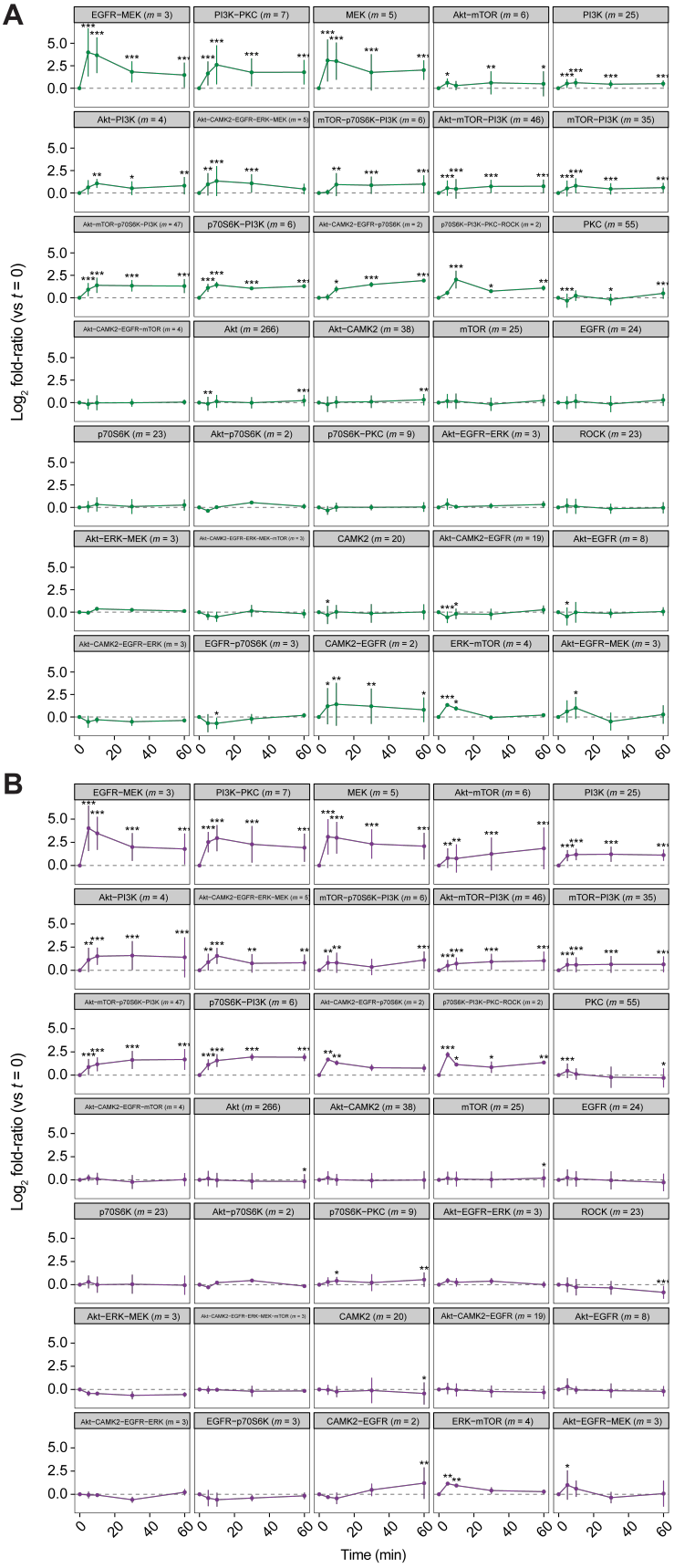


Figure 5.4: Branches of the defined network show substantial plasticity with EGF and IGF-1 stimulation. (A) Lineplots showing the temporal profiles of the CTAM groups with EGF stimulation (with $m \geq 2$). (B) Temporal profiles of CTAM group activities with IGF-1 stimulation (with $m \geq 2$). Data points represent the mean $\pm$ SD. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. P-values determined from a Z-test as outlined in Fig 5.1.

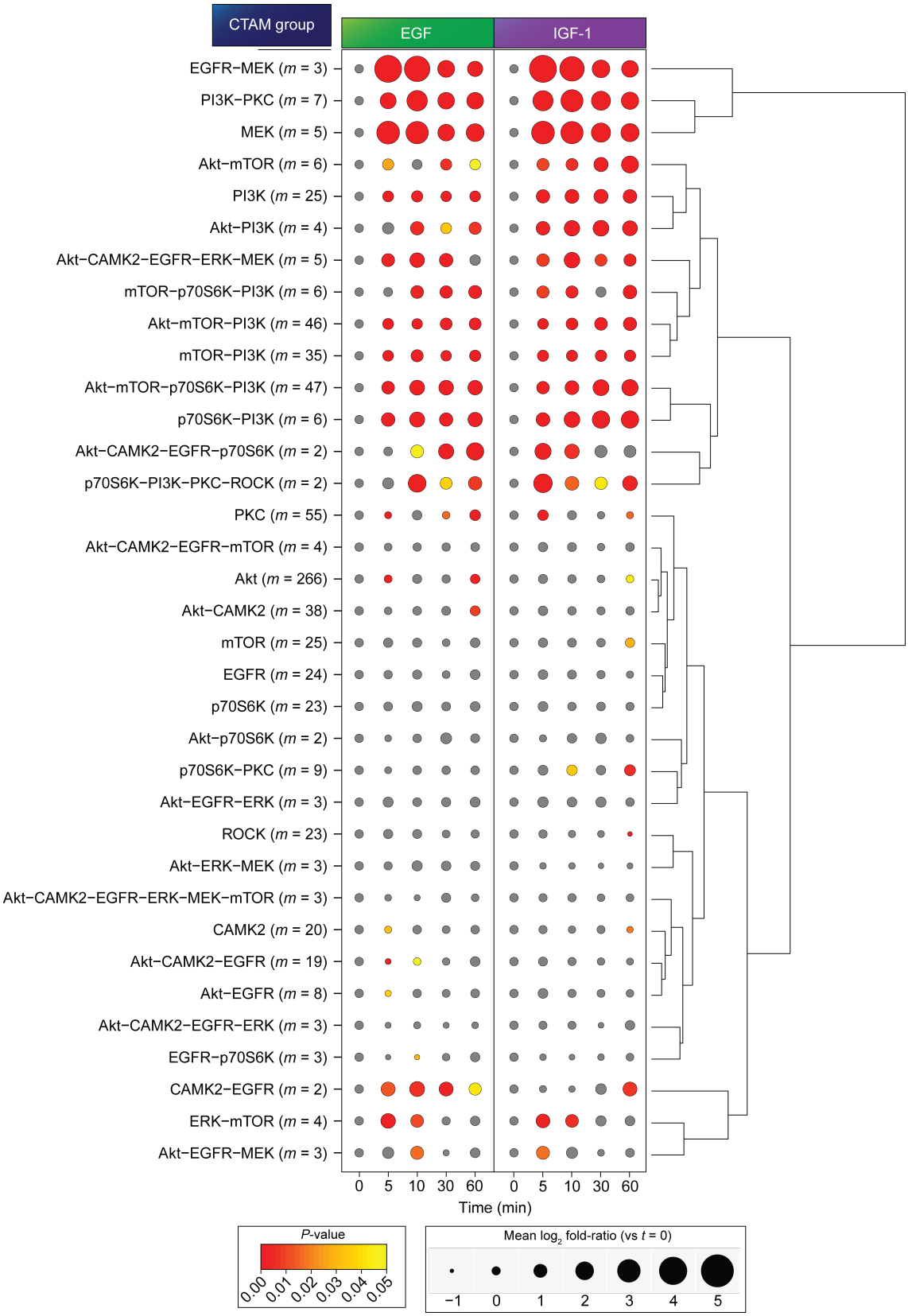


Figure 5.5: CTAM group profiling allows the capture of network activity at each time-point. Dotplot representing the temporal profiles of CTAM groups upon EGF and IGF-1 stimulation, as indicated. Each dot represents the mean $\log_2$ fold-ratio (vs $t = 0$ min) for the relevant CTAM group's constituent phosphopeptides, for each condition. The colour of each point represents the statistical significance associated with the change in abundance. The dendrogram represents the unsupervised, hierarchical clustering of the $\log_2$ fold-ratios for each CTAM group (Euclidean distance metric). *P*-values determined from a *Z*-test as outlined in Fig 5.1.

more sustained impact on PI3K/Akt signalling (Fig 5.2-5.4).

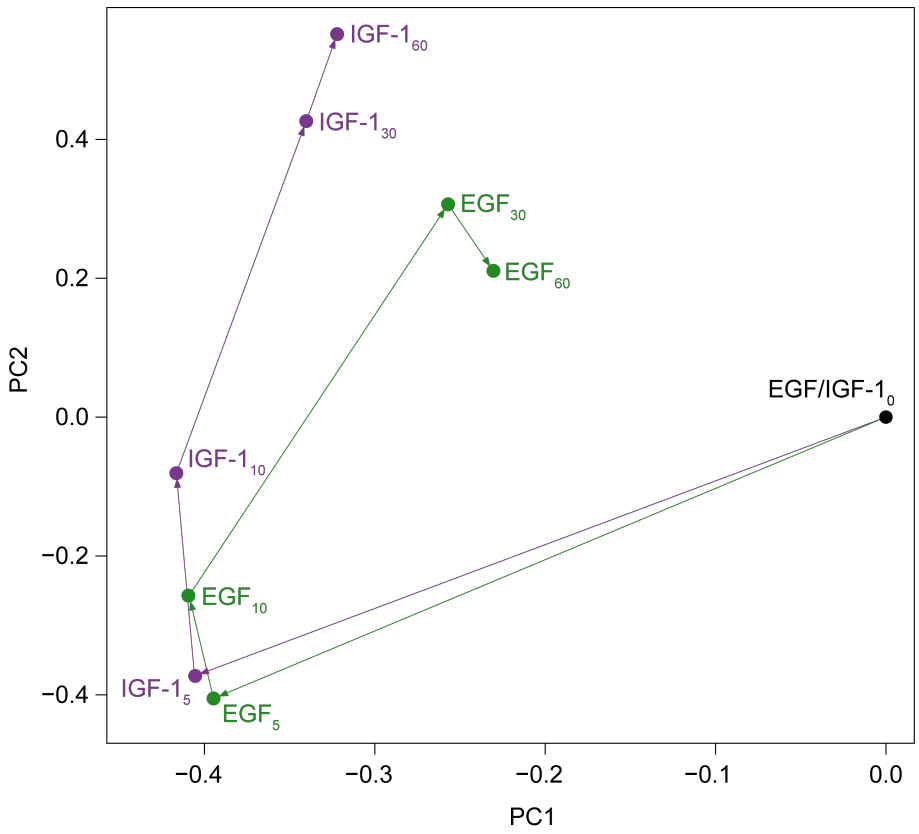


Figure 5.6: Principal component analysis of the CTAM network data resolves the late time-point differences in signalling between EGF and IGF-1. PCA of the time-course data as shown in Fig 5.5. Each point summarises the behaviour of each of the CTAM groups at the relevant time-point. Points are coloured according to the relevant growth factor (EGF, green; IGF-1, purple).

5.3.2 Intra-CTAM group homogeneity

Alongside the observation that the CTAM groups accurately represent the signalling within a system under well-known conditions, it was hypothesised that if CTAM groups truly represented the functional activation of the kinases and pathways they represent, the individual members of each group should demonstrate similar time-dependent behaviour. Although the error associated with each calculated mean is displayed in Fig 5.4 (one standard deviation), to assess each CTAM group's variability in a systematic manner, the coefficient of variations (CVs) for each CTAM group, at each time-point, and for each growth factor, were calculated. Plotting the distributions of the resulting values (Fig 5.7) revealed that the median of the aggregated CVs within each CTAM group, at each time-point, was 0.454 (45%) and 0.518 (52%) for EGF and IGF-1 respectively. These values, alongside the fact that the majority of the data-points were distributed between 0-1.0 (0-100%) suggested that the individual phosphopeptides within each CTAM group did indeed follow similar patterns of abundance as a function of time.

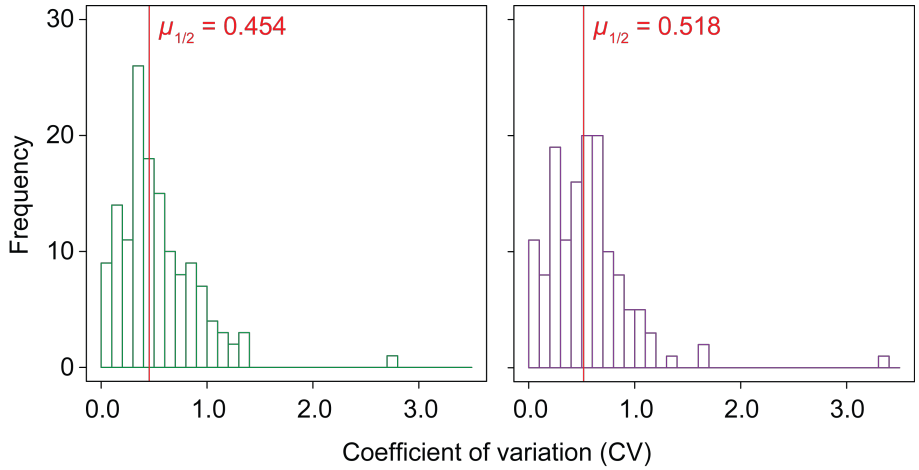


Figure 5.7: Analysis of the variability in CTAM group behaviour reveals intra-group homogeneity in responses. Histograms representing the distribution of coefficient of variations for each CTAM group under stimulation with each growth-factor. Red lines represent the median ($\mu_{1/2}$) in each case. The histograms are coloured according to which growth factor they represent (EGF, green; IGF-1, purple).

This homogeneity of response is visualised for two CTAM groups, one with a large number of members (Akt-mTOR-PI3K) and one with fewer (EGFR-MEK), in Fig 5.8. These data lend support to the notion that the branches of the defined network did in fact represent the functional activation of the kinases they represent, as the substrates – albeit with a degree of variability – showed the same trends upon growth factor stimulation. This observation held true in both a highly and sparsely populated CTAM group (Fig 5.8A and B respectively).

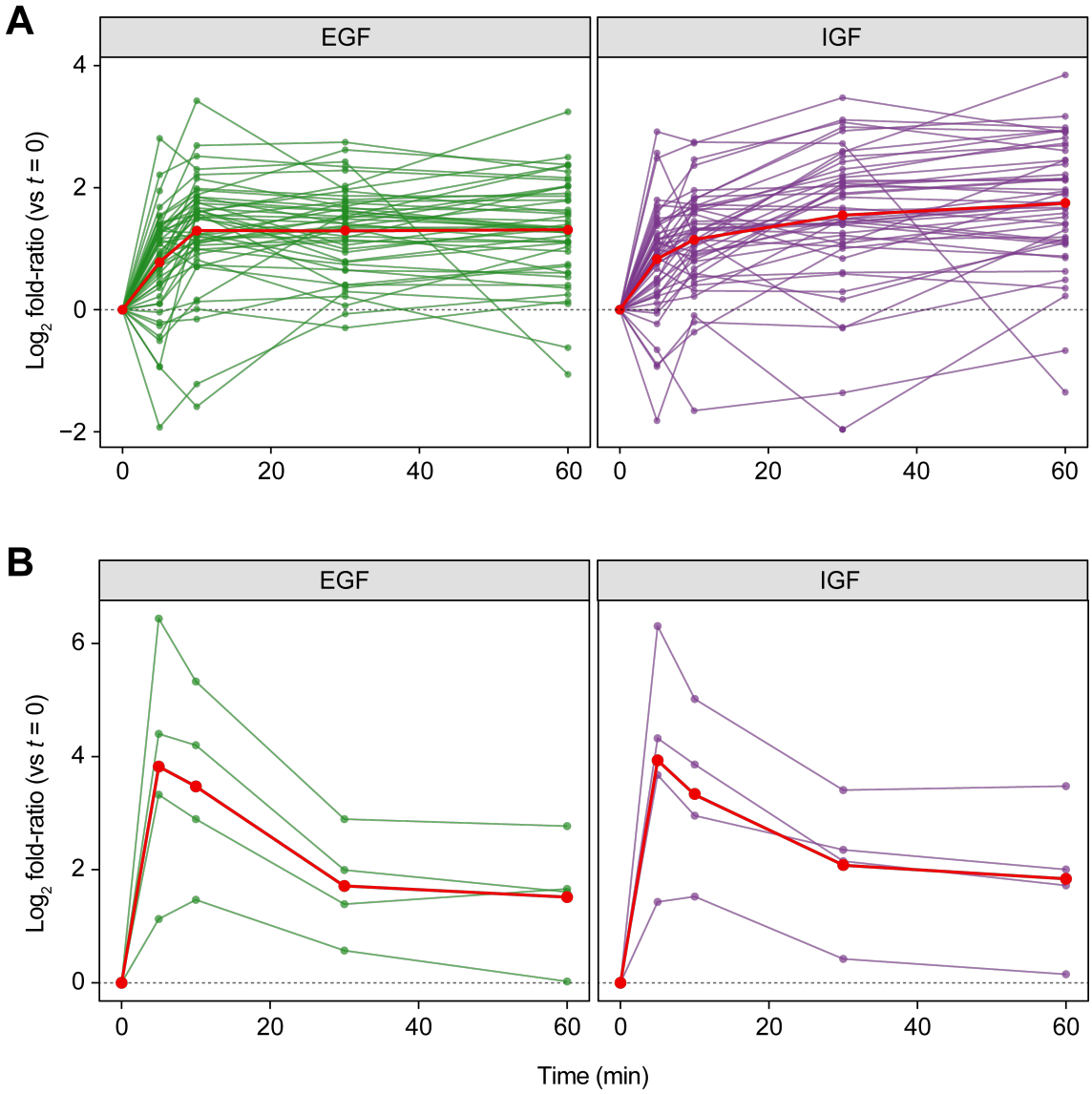


Figure 5.8: Examples of homogeneity in CTAM response to EGF and IGF-1. (A) Lineplots showing the temporal profiles of the individual CTAMs within the Akt-mTOR-PI3K group ($m = 46$) with EGF (green) and IGF-1 stimulation (purple). (B) Temporal profiles of individual CTAMs within the EGFR-MEK group ($m = 4$) with EGF and IGF-1 treatment. Red data points represent the mean of the green/purple data-points at each time-point.

5.4 Conclusions

A strategy was designed to test the functional relevance of the CTAM groups (and the network they form) defined in Chapter 4, and to test whether it was possible to assess network activity in an untargeted and global manner (Fig 5.1). Accordingly, the quantitative dynamics and plasticity of this network under EGF and IGF-1 stimulation were assessed through the use of label-free phosphoproteomics.

MCF7 cells were serum-starved and treated with 100 ng·mL^{-1} of either EGF or IGF-1 at five different time-points. The phosphoproteome of these cells was then analysed by nLC-MS/MS. Applying a robust statistical framework to the resulting data allowed the extraction of the magnitude and significance of changes in phosphopeptide abundance compared to the cells in a starved state (Table 5.1). Analysis of known downstream markers of EGFR and IGF-1R ligand stimulation within the MS dataset revealed that the MS data were able to recapitulate the known effects of these growth factors on the phosphoproteome (Fig 5.2). Western blots for canonical activatory phosphorylation sites on Akt and ERK1/2 served to validate that the MS data provided a faithful representation of the signalling events occurring in these cells upon stimulation (Fig 5.3). These data also confirmed the differential effects of EGF and IGF-1 on Akt phosphorylation (and therefore potential activation) observed in the MS data (Figs 5.2 and 5.3).

Following this validation, the quantitative dynamics of the defined CTAM grou-ps were assessed systematically with in-house developed scripts (as outlined in Fig 5.1). Firstly, the CTAMs identified in the experiment were extracted from the data. CTAM groups with $\geq$ 2 CTAMs identified (i.e. m) were then taken forward for further analysis. The mean $\log_2$ fold-ratio of each CTAM group (and associated error) was then calculated for each time-point. The statisical significance of changes in these CTAM groups was assessed by means of a modified Z-test (Fig 5.1 and ref[16]). This analysis revealed that many branches of the network exhibited plasticity under stimulatory conditions (Fig 5.4). Consistent with previous studies, this analysis showed that the network underwent acute and short-lived increases in MAPK-related kinase activities (EGFR-MEK and MEK). Furthermore, these data showed that the network underwent smaller but more sustained increases in Akt-mTOR-PI3K pathway activities. Clustering of these data revealed the network branches that showed similar kinetics of activation/inhibition over the time-course, whilst also revealing the differences between the responses to the two growth factors (Fig 5.5). In-line with the data shown in Figs 5.2 and 5.3, IGF-1 appeared to have a larger and more sustained impact on Akt-mTOR-PI3K signalling than EGF.

The difference in response of the cells to the two growth factors was assessed further using a principal component analysis. This showed that, alongside the patterns observed in Figs 5.2-5.5, IGF-1 had a more sustained impact on the phosphoproteome, as the state of the network at 60 minutes after IGF-1 stimulation was further away from the basal state (in PC-space) than the same time-point with respects to EGF. This revealed that this method of CTAM/network analysis was not only able to recapitulate the known effects of the growth factors on the phosphoproteome of MCF7 cells, but also resolve differences between them and provide insight into both EGF and IGF-1 signalling.

It was postulated that, should the CTAM groups/network branches represent functionally relevant signalling pathways then, alongside their accurate representation of signalling, the individual members within these groups should demonstrate similar temporal behaviour. To assess this, the CVs for each of the CTAM groups at each time-point were calculated and their distributions under both EGF and IGF-1 stimulation analysed (Fig 5.7). This analysis showed that the medians of these distributions were 45% and 52% for EGF and IGF-1 respectively, suggesting that the

individual CTAMs did indeed demonstrate intra-group homogeneity in their responses. This was assessed visually for two CTAM groups; one with a large number of members (Akt-mTOR-PI3K, m = 46), and one more sparsely populated (EGFR-MEK, m = 4) (Fig 5.8). This demonstrated the level of homogeneity in response of the individual CTAMs, as the vast majority in both the Akt-mTOR-PI3K and EGFR-MEK groups followed similar temporal profiles.

Collectively, these data suggested that the defined network, and the conceptual methodology by which its activity was measured, allowed the reconstruction of the effects on signalling caused by EGF and IGF-1 that was consistent with both previous studies and validation with an orthogonal method. Theoretically then, the defined network could be utilised to accurately assess kinase signalling in MCF7 cells, without prior knowledge as to the expected effects of the relevant treatment(s).

Chapter 6

Signalling network plasticity in models of acquired resistance to chronic PI3K or mTORC1/2 inhibition

6.1 Introduction and aims of the study

The results presented in Chapter 5 detailed the utility of the defined network, alongside a computational method, for the accurate assessment of the activity of the kinases within the network (and the connections between them) in an unbiased manner. It was then reasoned that, to investigate the effects of chronic inhibition of PI3K and mTORC1/2 and the subsequent acquisition of resistance, this method would reveal how the network is remodelled under these conditions. Importantly, this method would be without bias towards a particular signalling node.

The acquisition of resistance to targeted therapies, as discussed in Chapter 1, can occur through numerous mechanisms. These often include point-mutations on the target kinase(s), resulting in an abrogation of drug affinity whilst maintaining kinase activity. In recent years, however, numerous studies have demonstrated the ability of the targeted signalling pathway to compensate for the chronic inhibition of a particular signalling node by up-regulating or modifying downstream or upstream pathway members[106–108,110,112]. Alongside these phenomena, it has also been shown that pathways parallel to that being targeted can also compensate for the loss its activity; often through the up-regulation of an alternate RTK or cytoplasmic kinase[113,119]. Despite the uncovering of these mechanisms, it has become increasingly clear that signalling pathways are not as linear and monodirectional as they are often conceptualised to be. As a result, kinase signalling has become increasingly appreciated to involve large, highly interconnected networks of kinases and their substrates, which act in a concerted manner to translate extracellular stimuli into appropriate responses. The remodelling and plasticity of these networks in the context of chronic kinase inhibition and acquired resistance has not been studied in depth; however, mass spectrometric technology has rendered questions of this nature more tractable and less biased towards particularly well-characterised signalling nodes.

The objective of the experiments presented in this Chapter was to investigate the plasticity of the signalling network in models of acquired resistance to inhibitors of both PI3K and mTORC1/2. The ultimate aim of this was to increase our understanding of how kinase signalling changes in response to chronic inhibition of particular network nodes. To achieve this, a single MCF7 culture (at low passage number) was homogenised and split into six identical populations of cells. Three of these new cultures were treated with a gradually increasing concentration of GDC-0941 (pan-class I PI3K inhibitor) and three with KU-0063794 (mTORC1/2 inhibitor). The initial concentration of each drug was 100 nM and this was raised by 100 nM·passage^{-1} to a maximum of 1 μM (Fig 6.1).

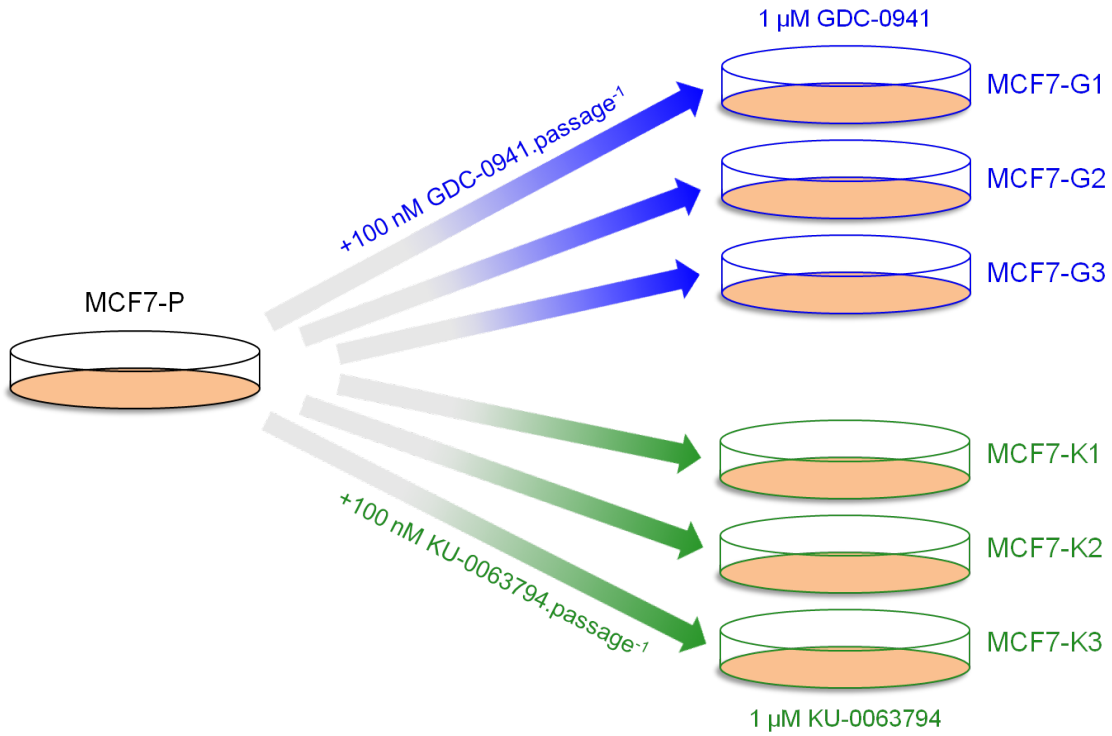


Figure 6.1: Generation of models of acquired resistance through the chronic treatment of identical MCF7 cultures with either GDC-0941 or KU-0063794. Parental MCF7 cells (MCF7-P) were split into six identical cultures. Three were treated with 100 nM GDC-0941 (MCF7-G1:3, blue) and three with 100 nM KU-0063794 (MCF7-K1:3, green). The concentration of the drug was increased by 100 nM at each passage of the cells, until a final concentration of 1 μM. Each cell-line was treated identically throughout and the three cultures treated with each drug were maintained at the same passage number throughout the treatments and subsequent experiments.

6.2 Defining the sensitivity of parental cells

Prior to the creation of the resistant cell-lines, the sensitivity of the MCF7-P (parental) cells to the two compounds was determined through the use of an 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS) assay. The cells were treated with either compound on a $\log_{10}$ scale between 0.001 and 10 μM for 48 hours. The reduction of the MTS compound through mitochondrial action was then used a proxy for cell viability. The experiment was performed in quadruplicate (Fig 6.2).

The data shown in Fig 6.2 demonstrate that the parental cells were initially sensitive to both compounds. Fitting third-order polynomial models to the data and solving the relevant equations, however, showed that the parental cells were more sensitive to GDC-0941 ($IC_{50} \approx 162$ nM) than to KU-0063794 ($IC_{50} \approx 503$ nM).

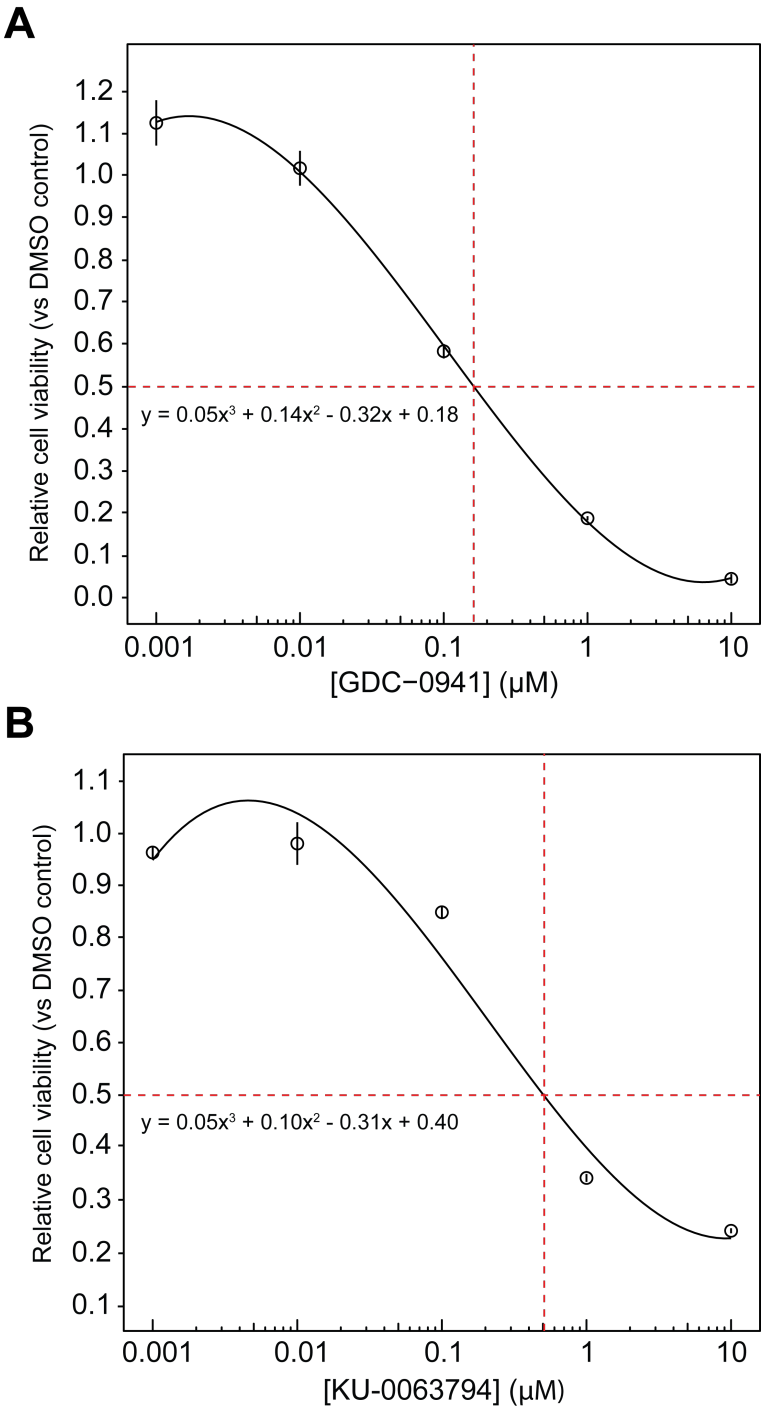


Figure 6.2: The parental cells are sensitive to the compound at nM concentrations. (A) Viability of the parental cells under treatment with GDC-0941. (B) Viability of the parental cells under treatment with KU-0063794. Data-points represent the mean ± SEM (n = 4) relative to DMSO control. Solid lines represent third-order polynomials fitted to the data. Dashed lines represent the computed IC_{50} of the relevant compound.

6.3 Profiling of the resistant cell-lines

6.3.1 Determination of acquired resistance

After having grown the six cell-lines up to a maximum of 1 μM of the relevant compound, an experiment was designed to determine whether the cells were indeed resistant to the anti-proliferative effects of the drugs. To this end, the proliferation in the presence of each drug was compared between the resistant and parental cell-lines (Fig 6.3). Cells were treated with 1 μM of the relevant inhibitor and cultured for 0, 1, 3, or 5 days. The number of viable cells at each time-point was then measured by means of a Beckman Coulter ViCELL counter as per the manufacturer's instructions.

These data served to demonstrate that the resistant cell-lines were indeed resistant to the compounds they had been treated with, as they were able to proliferate in the presence of drug where the parental cells were unable to do so (Fig 6.3A-B). Of note, however, the rates of proliferation were not identical for each cell-line resistant to same drug. For example, MCF7-K1 and K3, despite still being able to proliferate in the presence of KU-0063794, grew much more slowly than MCF7-K2 (Fig 6.3B).

The resistance of each of the cell-lines was also assessed by means of the MTS and crystal violet (CV) cell viability assays (Fig 6.3C-D). These experiments, although performed after only 2 days of treatment, also demonstrated that the cells were able to proliferate in the presence of the relevant compound where the parental cells were not. The differences in proliferation rates between the resistant cells were also observed with these techniques, providing further evidence to suggest that these differences were robust.

Overall, these data showed that the resistant cell-lines were resistant to their respective compounds; however, the cells treated with GDC-0941 were comparatively more resistant than those treated with KU-0063794 – this being in-line with the observation that the parental cells were initially less sensitive to treatment with KU-0063794 (Fig 6.2).

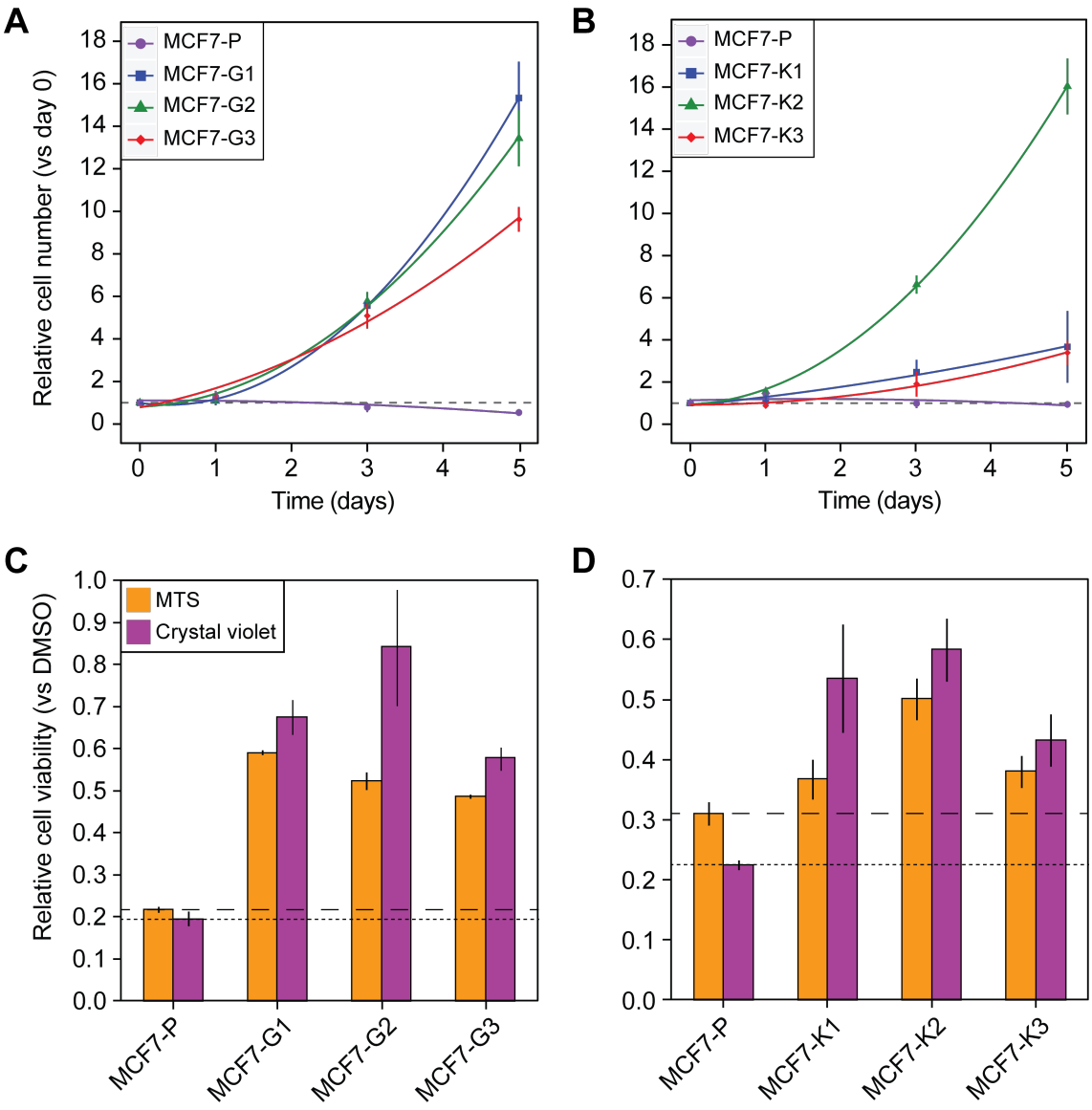


Figure 6.3: The resistant cell-lines are able to proliferate in the presence of their respective inhibitors. (A) Proliferation of the parental and GDC-0941-resistant MCF7 cells in the presence of 1 μM GDC-0941. (B) Proliferation of the parental and KU-0063794-resistant MCF7 cells in the presence of 1 μM KU-0063794. Data-points represent the mean $\pm$ SD ($n = 3$). Solid lines represent second order polynomials fitted to the data and are coloured according to the cell-line they represent. (C) MTS and CV cell viability assays for the parental and GDC-0941-resistant MCF7 cells treated with 1 μM GDC-0941 for 48 hours. (D) MTS and CV cell viability assays for the parental and KU-0063794-resistant MCF7 cells treated with 1 μM KU-0063794 for 48 hours. Bars represent the mean $\pm$ SD ($n = 3$). Long-dashed and short-dashed lines represent the parental cell-line's proliferation as determined by the MTS and CV assays respectively.

6.3.2 Phosphoproteomic and proteomic profiling of the resistant and pare-ntal models

Having established that the resistant cell-lines were resistant to the anti-proliferative effects of the two compounds, the changes in the signalling network between the parental and resistant lines could be compared. To accomplish this, each of the cell-lines was grown in duplicate and their phosphoproteomes analysed by nLC-MS/MS. For the purposes of this experiment, the parental cells were grown in 0.1% DMSO and the resistant cell-lines maintained in 1 μM of the relevant inhibitor. Due to the disparate rates of the growth of the resistant cell-lines, the MCF7-G and MCF7-K cells were processed separately, both with a duplicate MCF7-P control. To control for changes in expression of the unmodified proteins represented by the identified phosphorylation sites, the total proteome of these cells was also analysed. These measurements were then used to normalise the phosphoproteomics data and thus calculate stoichiometry for each phosphopeptide (where these data were available). A quality control assessment of these data is shown in Appendix 4.

The quantitative data, as with the experiments presented in Chapters 4 and 5, were analysed using *limma*. The linear model designs for the phosphoproteomics and proteomics data included a single MCF7-P control factor, a factor for each resistant cell-line, and a factor for the different experiments. The correlations between the technical triplicates (phosphoproteomics) and duplicates (proteomics) were accounted for in the design. The results of this differential abundance analysis are shown in Tables 6.1 and 6.2.

Cell-line	**FR $<$ 0; adjusted *P* $\leq$ 0.05**	**FR $>$ 0; adjusted *P* $\leq$ 0.05**
MCF7-G1	1558	1486
MCF7-G2	1087	1151
MCF7-G3	1040	1070
MCF7-K1	2084	1318
MCF7-K2	1887	1511
MCF7-K3	2146	1750

Table 6.1: Differential abundance analysis of the phosphoproteomics data using *limma*. The number of phosphopeptides either significantly increased or decreased in abundance compared to parental cells are shown (n_{total} = 11,307). FR, $\log_2$ fold-ratio vs DMSO control; adjusted *P*, Benjamini-Hochberg multiple testing corrected *P*-value derived from a moderated *t*-test.

Cell-line	**FR $<$ 0; adjusted *P* $\leq$ 0.05**	**FR $>$ 0; adjusted *P* $\leq$ 0.05**
MCF7-G1	24	49
MCF7-G2	185	82
MCF7-G3	195	323
MCF7-K1	18	25
MCF7-K2	73	90
MCF7-K3	40	61

Table 6.2: Differential abundance analysis of the proteomics data using *limma*. The number of proteins either significantly increased or decreased in abundance compared to parental cells are shown (n_{total} = 2,593). FR, $\log_2$ fold-ratio vs DMSO control; adjusted *P*, Benjamini-Hochberg multiple testing corrected *P*-value derived from a moderated *t*-test.

The results shown in Tables 6.1 and 6.2 highlight that a large number of significant differences, at both the phosphopeptide and protein levels, existed between the resistant and parental cells. Interestingly, the number of differentially regulated phosphopeptides and proteins was heterogeneous between each of the resistant cell-lines, suggesting that their phosphoproteome and proteome had been remodelled differently.

6.3.3 CTAM network analysis of the resistant cell-lines

Following on from the differential abundance analysis that demonstrated clear quantitative differences in phosphopeptide and protein abundances between the parental and resistant cell-lines, the quantitative data – in the form of computed stoichiometry ratios – were used to profile the CTAM network as discussed in Chapters 4 and 5 (Fig 5.1). The results of this analysis are shown in Fig 6.4.

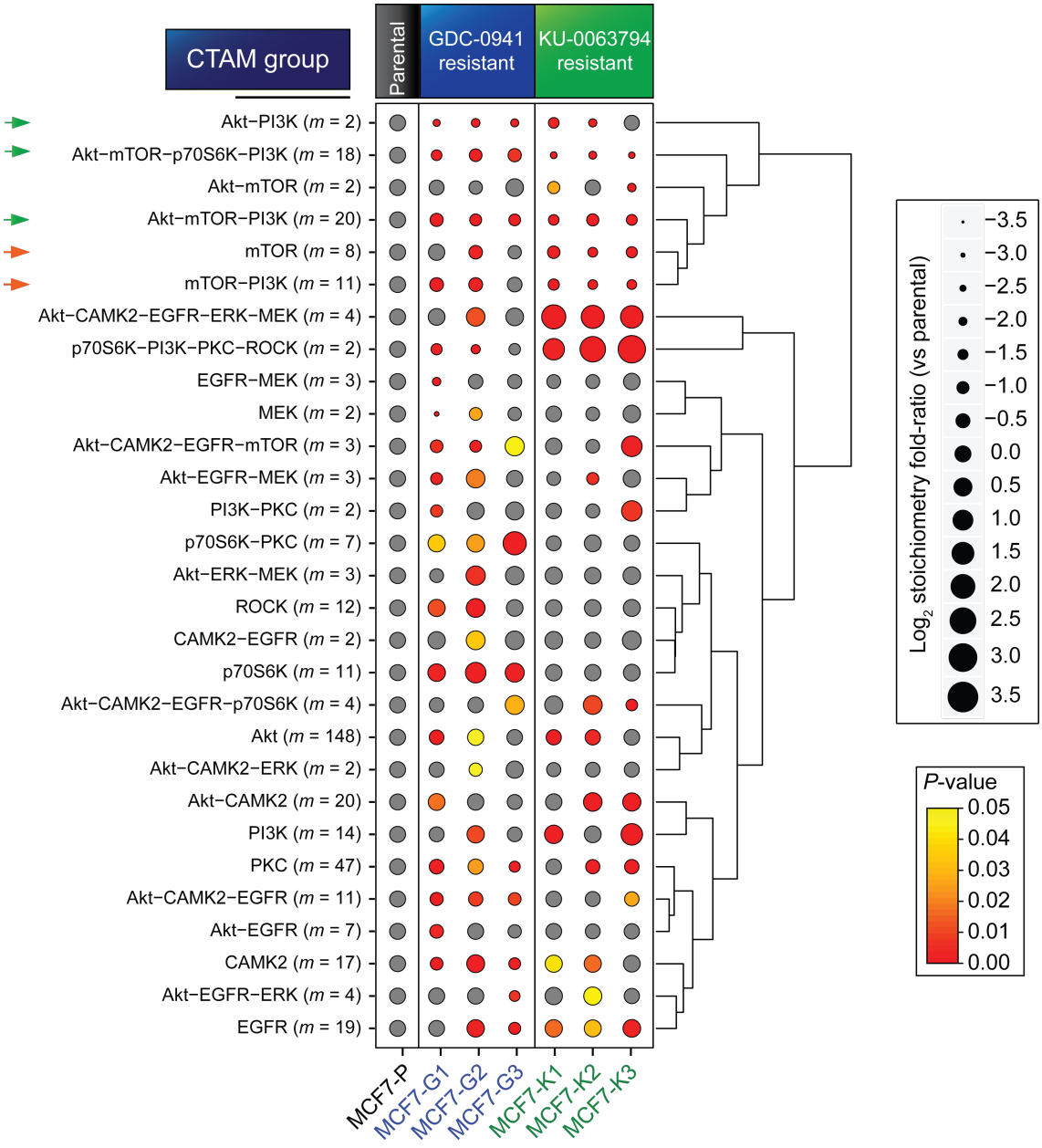


Figure 6.4: CTAM analysis of the resistant cell-lines reveals substantial remodelling of the signalling network. CTAM group/network branch profiles for each of the resistant cell-lines compared to the parental cells. The size of each dot represents the mean $\log_2$ fold-ratio of each CTAM group's constituent phosphopeptides relative to the parental cell-line, where each individual phosphopeptides abundance is normalised to the relative abundance of the unmodified protein. Colours represent the statistical significance of enrichment, as determined by a modified *Z*-test. The dendrogram represents the hierarchical clustering of the mean $\log_2$ fold-ratios (Euclidean distance metric).

Analysing the data in this manner revealed that the activity of the different branches (i.e. CTAM groups) of the defined signalling network were substantially altered in activity when compared to the parental cells. Importantly, as previously alluded to, these differences were independent of changes at the protein level. Hierarchical clustering of these data revealed the CTAM groups

which displayed similar quantitative behaviour across the cell-lines. This clustering showed that the majority of CTAM groups representing PI3K and mTOR were reduced in abundance in both the MCF7-G (Fig 6.4, green arrows) and MCF7-K (Fig 6.4, green and orange arrows) compared to the parental cells. This suggested that these kinases remained inhibited by each compound and that instead, compensatory mechanisms were allowing continued proliferation. In agreement with this hypothesis, a number of other CTAM groups were increased in abundance compared to the parental levels.

This analysis also highlighted that the responses of the network evolved in response to GDC-0941 (left, blue) were distinct to those for KU-0063794 (right, green). Moreover, a degree of heterogeneity between the individual cell-lines within each drug group was also observed. To assess the differences between this inter- and intra-group heterogeneity between the two groups of cell-lines more objectively, a principal component analysis was performed on these data (Fig 6.5).

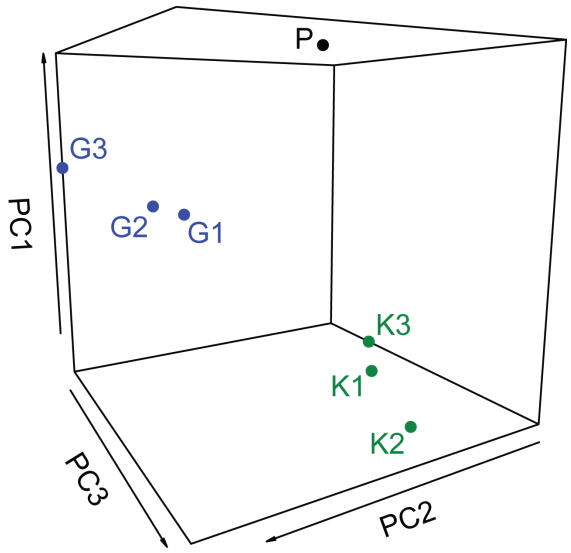


Figure 6.5: The resistant cell network profiles are distinct between MCF7-G and -K cells. Scatter plot representing the first three principal components of the network data shown in Fig 6.4. GDC-0941-resistant cells, blue; KU-0063794-resistant cells, green; parental cells, black; PC, principal component.

The analysis shown in Fig 6.5 reinforced the idea that the activity of the network within the resistant cells had evolved to become substantially different to the parental cells, as the MCF7-G and -K cells separated from MCF7-P cells in PC-space. This analysis also showed that the responses evolved in the MCF7-G cells were distinct to those in the MCF7-K cells, as these also separated well. Most strikingly, a degree of separation was observed between the individual cell-lines within each group (e.g. MCF7-G1, -G2 and -G3), mirroring the heterogeneity of CTAM group activation noted in Fig 6.4 and suggesting that even cells resistant to the same compound remodelled their network in subtly different manners.

6.3.4 Validation of the MS and CTAM analyses

To provide further evidence for the maintained reduction of PI3K and mTOR signalling in the resistant cells, a number of known downstream phosphorylation sites of these kinases were extracted from the data (Fig 6.6A). In addition, Western blots were performed for known activatory phosphorylation sites on kinases downstream of PI3K and mTOR (Fig 6.6B). Consistent with the CTAM analysis shown in Fig 6.5, a number of known mTOR substrates were significantly reduced in abundance in both the MCF7-G and -K cells. In-line with the patterns observed in these data and those shown in Figs 6.4 and 6.5, Western blots against Akt Ser^{473} and p70S6K (RPS6KB1) Thr^{389} demonstrated a substantial reduction of Akt and p70S6K phosphorylation in these cells, suggesting that the signalling of these kinases was indeed reduced. These data also mirrored the patterns observed in the MS data (Fig 6.6A) as the down-regulation of mTOR signalling (based on the phosphorylation of p70S6K) was more substantially reduced in the MCF7-K than MCF7-G cells.

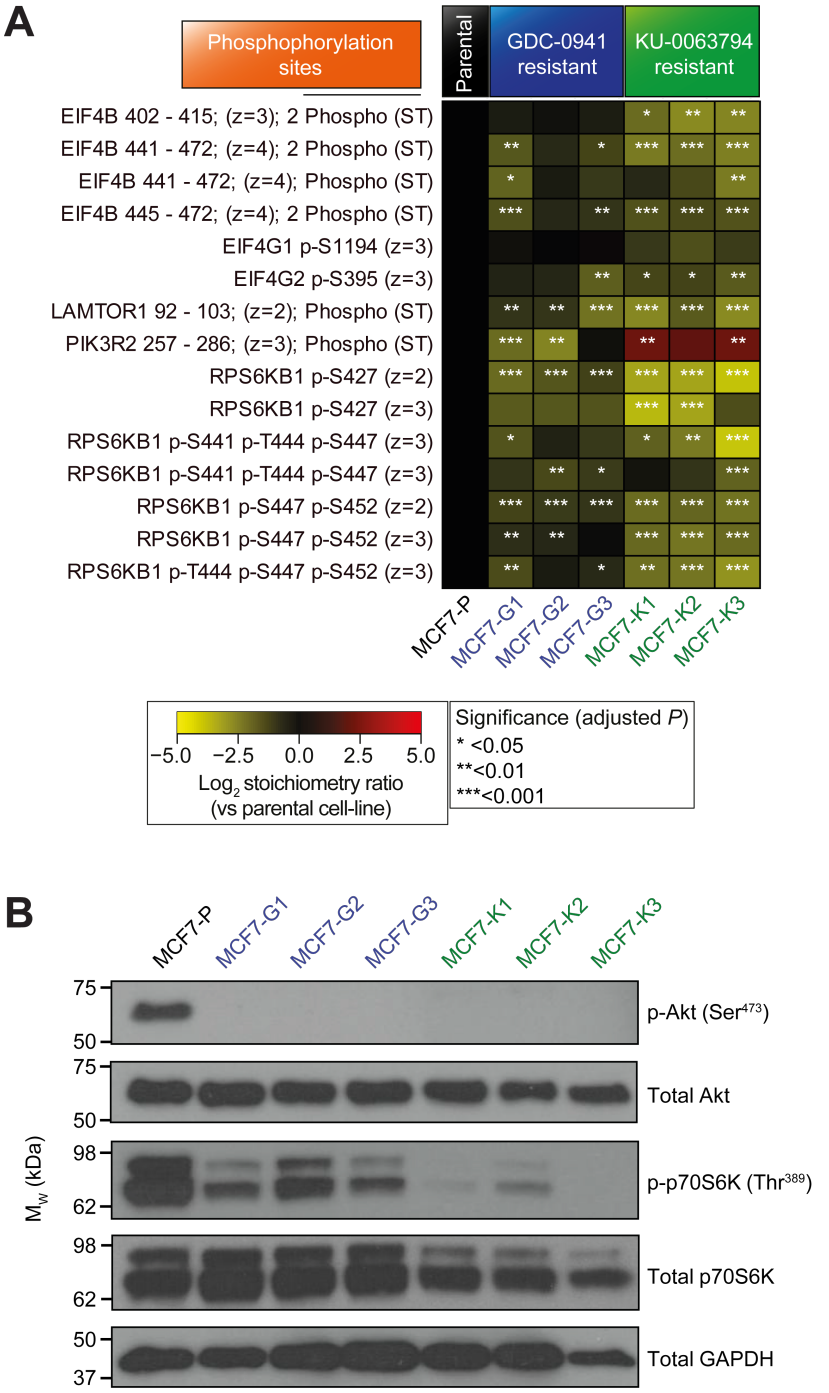


Figure 6.6: PI3K and mTOR signalling is reduced in the resistant cell-lines. (A) Heat map of selected phosphorylation sites representing the activity of PI3K and mTOR signalling in the resistant cells (where proteomics data was available to provide stoichiometry). Uniprot protein ID (gene name), phosphorylation sites, peptide charge, and other peptide modifications are shown. If the δ-score for the peptide is < 10, the peptide is shown in the following format: "protein gene name (sequence residues); charge; modification". (B) Westerns blots for known markers of PI3K/mTOR signalling activity in the resistant and parental cells. Abundance of the unmodified protein is shown in each case.

6.4 Linking network activities to phenotypic outputs

6.4.1 Responses of resistant and parental cells to a panel of kinase inhibitors

To investigate whether remodelling of the network had functional consequences for the resistant cells, and that the observed intra- and inter-group heterogeneity could be recapitulated with an orthogonal method, the responses of the cells to a panel of small-molecule inhibitors was assessed using both the MTS and CV assays. It was hypothesised that, as the response to kinase inhibitors is a function of signalling network activity, the resistant cells should respond differently to both parental cells and one another. An unbiased, multivariate analysis (PCA) should then be able to resolve these differences and reaffirm the observations made in Figs 6.4 and 6.5. For these experiments, each of the cell-lines was plated in triplicate and treated with inhibitors of Akt, MEK, EGFR and CAMK2, both alone and in combination with GDC-0941 or KU-0063794. The resulting viability data from these experiments (normalised to DMSO control) is shown in Fig 6.7.

These experiments clearly demonstrated that, based on the differences between parental and resistant cells (dashed and coloured bars), the network remodelling observed in the phosphoproteomics and network analyses did indeed have an effect on the way in which these cells responded to other small-molecule kinase inhibitors. Moreover, these data recaptured the differences observed in the network between the MCF7-G and -K cells, as these responded differently to a number of different drug combinations. The MCF7-G cells, for example, although clearly resistant to GDC-0941, remained as sensitive to mTOR inhibition as the parental cells. On the contrary, the MCF7-K cells, alongside being more resistant to KU-0063794 than the parental cells, were also able to proliferate in the presence of 1 μM GDC-0941 (Fig 6.7). In-line with the observation that the network remodelling was heterogeneous between cells resistant to the same inhibitor, Fig 6.7 shows that the responses to the inhibitor panel were also subtly heterogeneous between individual cell-lines. MCF7-K1 and K3 cells were sensitive to the combined inhibition of PI3K/CAMK2 and PI3K/MEK, whereas the MCF7-K2 cells were able to proliferate more successfully in the presence of these drugs, for example (Fig 6.7).

To examine the relationships between each of the cell-lines based on the viability data presented in Fig 6.7, a principal component analysis was performed on this dataset (Fig 6.8). This analysis showed that the two sets of viability data, alongside the CTAM network analysis, were able to separate the MCF7-G and -K cells from the parental cell-line, providing further evidence to suggest that the network had been remodelled upon acquisition of resistance. The analyses were also able to discriminate between the MCF7-K and -G cells, further suggesting that the evolved responses to the two drugs were distinct. Alongside these observations – reflecting what was observed in the MS data – these analyses separated cell-lines resistant to the same drug, thus lending support to the notion that cells treated with the same compound, under the same conditions, remodelled their network in different ways to overcome chronic inhibition.

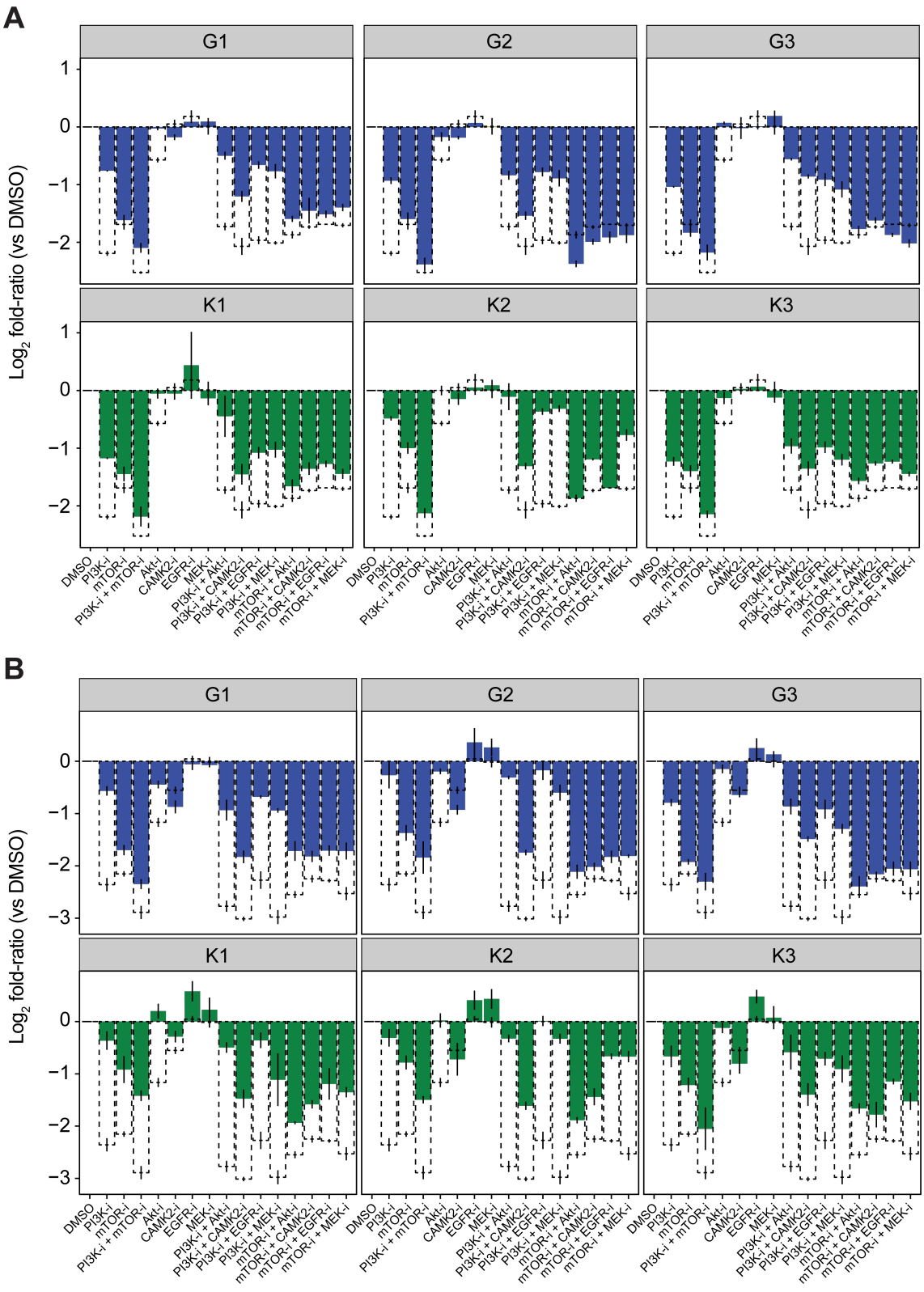


Figure 6.7: Network remodelling in resistant cells results in differential responses to small-molecule kinase inhibitors. (A) MTS cell viability data for the resistant and parental cell-lines treated with a panel of small-molecule kinase inhibitors (normalised to DMSO control). (B) CV cell viability for the identical experiment as in (A). Data points represent the mean $\pm$ SD ($n = 3$). Dotted lines represent the responses of the parental cells in each case. Akt-i, MK-2206 (1 μM); CAMK2-i, KN-93 (5 μM); EGFR-i, PD-153035 (1 μM); MEK-i, GSK-1120212 (0.5 μM); mTOR-i, KU-0063794 (1 μM); PI3K-i, GDC-0941 (1 μM).

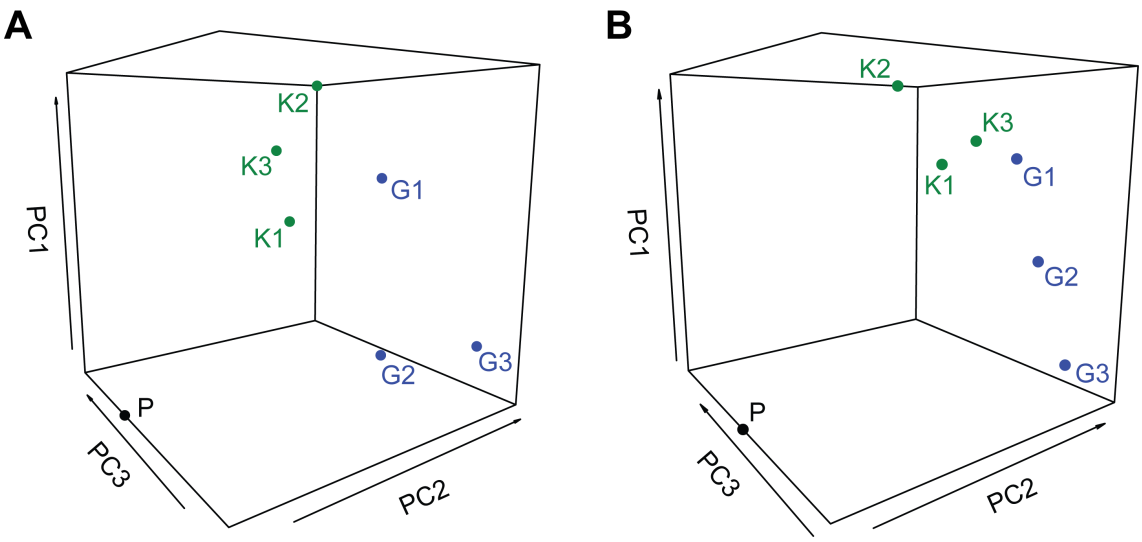


Figure 6.8: Multivariate analysis of the cell viability data separates the MCF7-G and -K cells. (A) Scatter plot of the first three principal components of the MTS cell viability data as in Fig 6.6A. (B) PCA of the CV cell viability data as in Fig 6.6B. GDC-0941-resistant cells, blue; KU-0063794-resistant cells, green; PC, principal component.

6.5 Rewiring of the signalling network in MCF7-G cells

The previous experiments provided evidence to suggest that: (i) the network was remodelled following chronic kinase inhibition; (ii) this remodelling was different for cells treated with GDC-0941 and KU-0063794 and; (iii) this remodelling was subtly different between cells resistant to the same compound. Despite this, these data were unable to distinguish if this was merely due to differences in the extents of signalling pathway activation between the resistant and parental cell-lines or whether the network was indeed rewired, with signals being redirected through different paths in the network. To distinguish between these two possibilities, it was hypothesised that if the network was indeed rewired, differences should be observed between the growth-factor induced phosphorylation kinetics in the resistant and parental cell-lines. To test this, an experiment was designed whereby the three MCF7-G cell-lines and the parental cells were starved for 24h (in FBS and DMSO/drug-free DMEM) and subsequently treated with 100 $ng{\cdot}mL^{-1}$ EGF for 0, 30, 60, or 120 minutes. The phosphoproteome of these cells at each time-point was then analysed by nLC-MS/MS in triplicate and the experiment was repeated to produce biological duplicates, thus resulting in a total of six analytical replicates. A quality control analysis of this dataset is shown in Appendix 5.

The quantitative data were normalised to, and statistical significance computed against, the 0 minute time-point for each respective cell-line. The results of the differential abundance analysis using *limma* are shown in Table 6.3. As an overview, these data provided preliminary evidence for a difference in the way in which the GDC-0941-resistant cells responded to EGF compared to the parental cells. These data showed that a large number of phosphopeptides significantly increase in abundance following EGF stimulation of the parental cells. In contrast, much fewer changes were observed following EGF stimulation of the resistant cell-lines. In comparison to the parental and MCF7-G1 and -G2 cells, the G3 cell-line appeared to be the least responsive to EGF treatment.

Cell-line/time-point	**FR $<$ 0; adjusted *P* $\leq$ 0.05**	**FR $>$ 0; adjusted *P* $\leq$ 0.05**
MCF7-P (t = 30 min)	589	336
MCF7-P (t = 60 min)	457	621
MCF7-P (t = 120 min)	224	488
MCF7-G1 (t = 30 min)	159	272
MCF7-G1 (t = 60 min)	169	226
MCF7-G1 (t = 120 min)	121	186
MCF7-G2 (t = 30 min)	228	233
MCF7-G2 (t = 60 min)	178	182
MCF7-G2 (t = 120 min)	246	190
MCF7-G3 (t = 30 min)	46	150
MCF7-G3 (t = 60 min)	10	83
MCF7-G3 (t = 120 min)	64	31

Table 6.3: Differential abundance analysis of the resistant cell growth factor time-course phosphoproteomics data using *limma*. The number of phosphopeptides either significantly increased or decreased in abundance compared to parental cells are shown (n_{total} = 4,228). FR, $\log_2$ fold-ratio vs DMSO control; adjusted *P*, Benjamini-Hochberg multiple testing corrected *P*-value derived from a moderated *t*-test.

6.5.1 Differences in EGF-induced phosphorylation kinetics between MCF7-G and -P cells

To assess the differences in EGF-responsive phosphorylation site kinetics between the resistant and parental cells in more depth, the dataset was filtered to only include those sites that demonstrated a statistically significant change compared to t = 0 minutes, in at least one condition. The $\log_2$ fold-ratios of the resulting 2,322 phosphopeptides were then hierarchically clustered and

visualised (Fig 6.9A).

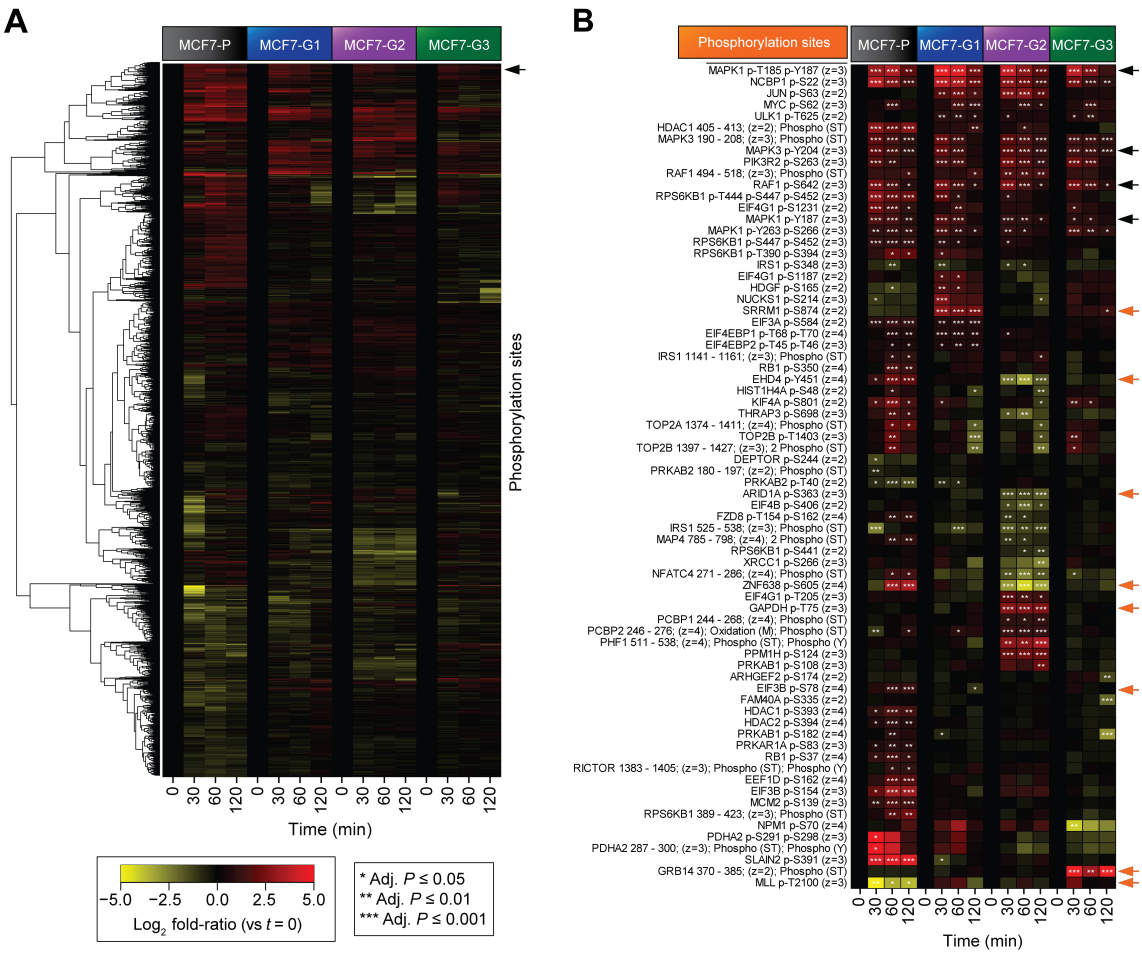


Figure 6.9: EGF-induced phosphorylation kinetics are distinct between MCF7-G and -P cells. (A) Heat map of the 2,322 significantly changing (adjusted $P \leq 0.05$) phosphopeptides for both the MCF7-G and -P cells treated with EGF for 0, 30, 60, and 120 minutes. The dendrogram represents the hierarchical clustering of each phosphopeptide based on their temporal behaviour (Euclidean distance metric). (B) Heat map of selected phosphorylation sites from the original 2,322, highlighting similarity in MAPK signalling (black arrows) and differences in signalling (orange arrows) between the parental and resistant cells. Uniprot protein ID (gene name), phosphorylation sites, peptide charge, and other peptide modifications are shown. If the δ-score for the peptide was < 10, the peptide is shown in the following format: "protein gene name (sequence residues); charge; modification".

The data shown in Fig 6.9A demonstrated that a core set of phosphorylation sites followed the same activation kinetics in the three resistant cell-lines and the parental cells (black arrow). Further analysis of this group revealed that these sites were closely associated with MAPK signalling (Fig 6.9B). The data also showed, however, that a large proportion of the assayed phosphopeptides demonstrated different kinetics between the resistant and parental cells and between the three resistant cell-lines (Fig 6.9B, orange arrows). This suggested that EGF-induced signals were flowing through different routes of the signalling network and therefore reinforced the view that the network was indeed rewired in these cells.

Analysing the CTAM signalling network in these cells also revealed similarities and differences in the way in which they responded to EGF (Fig 6.10). For example, as with the analysis shown in Fig 6.9, the resistant cells shared a marked activation of MAPK signalling upon EGF stimulation (Fig 6.10, black arrows). In addition, these cells appeared to still activate the Akt-mTOR-p70S6K-PI3K axis, alongside the parental cells (FIg 6.10, green arrow). In contrast to these observed similarities, several groups, although responsive in the parental cells, were non-responsive or even decreased in activity in the resistant cells (e.g. the Akt-CAMK2-EGFR and Akt-CAMK2-EGFR-ERK-MEK groups – Fig 6.10, orange arrows). To further distinguish between the response

of the phosphoproteome to EGF stimulation in these cells, multivariate analyses were performed on both the filtered phosphoproteomics data shown in Fig 6.9 and the CTAM network analysis of these data shown in Fig 6.10.

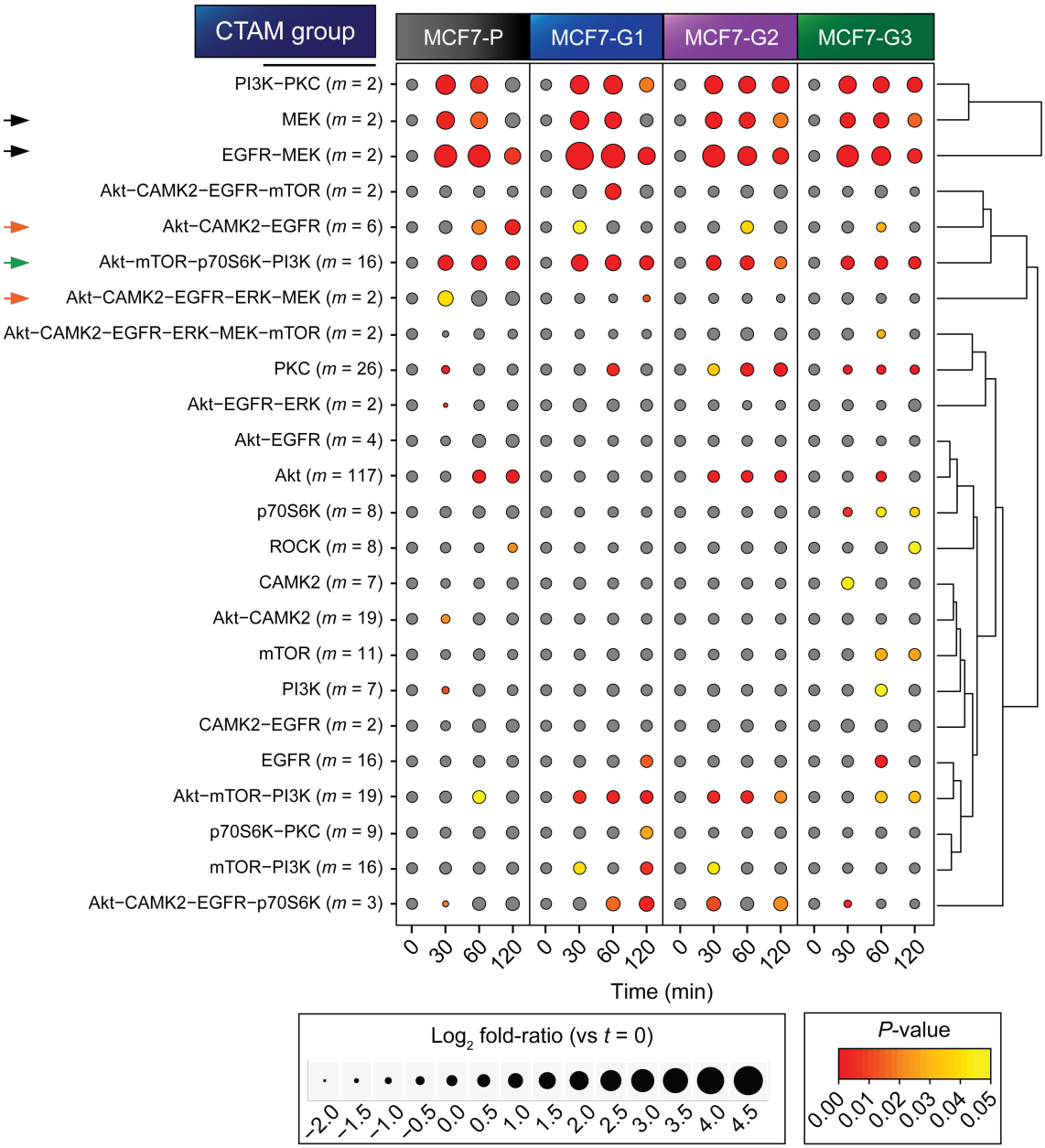


Figure 6.10: CTAM analysis further show that EGF-induced phosphorylation kinetics are distinct between MCF7-G and -P cells. CTAM network analysis of the resistant cell-line time-course experiment. The dendrogram represents the hierarchical clustering of each CTAM group based on their temporal behaviour (Euclidean distance metric).

The PCAs presented in Fig 6.11A and B demonstrate how the phosphoproteome/network was most different from the 0 minute time-point (i.e. a greater distance from these points in PC-space) at 30 minutes. This was a robust phenomenon, as it was observed in the responses of each of the cell-lines. These analyses also showed, however, that each of the resistant cell-lines' phosphoproteome 30 minutes following EGF stimulation, separated in PC-space from the parental cells and from one another (Fig 6.11A and B). Furthermore, the routes in PC-space by which the phosphoproteome/network relaxed back towards its basal state (0 minutes) were distinct for each of the cell-lines – this being observed with respect to both the filtered phosphoproteomics

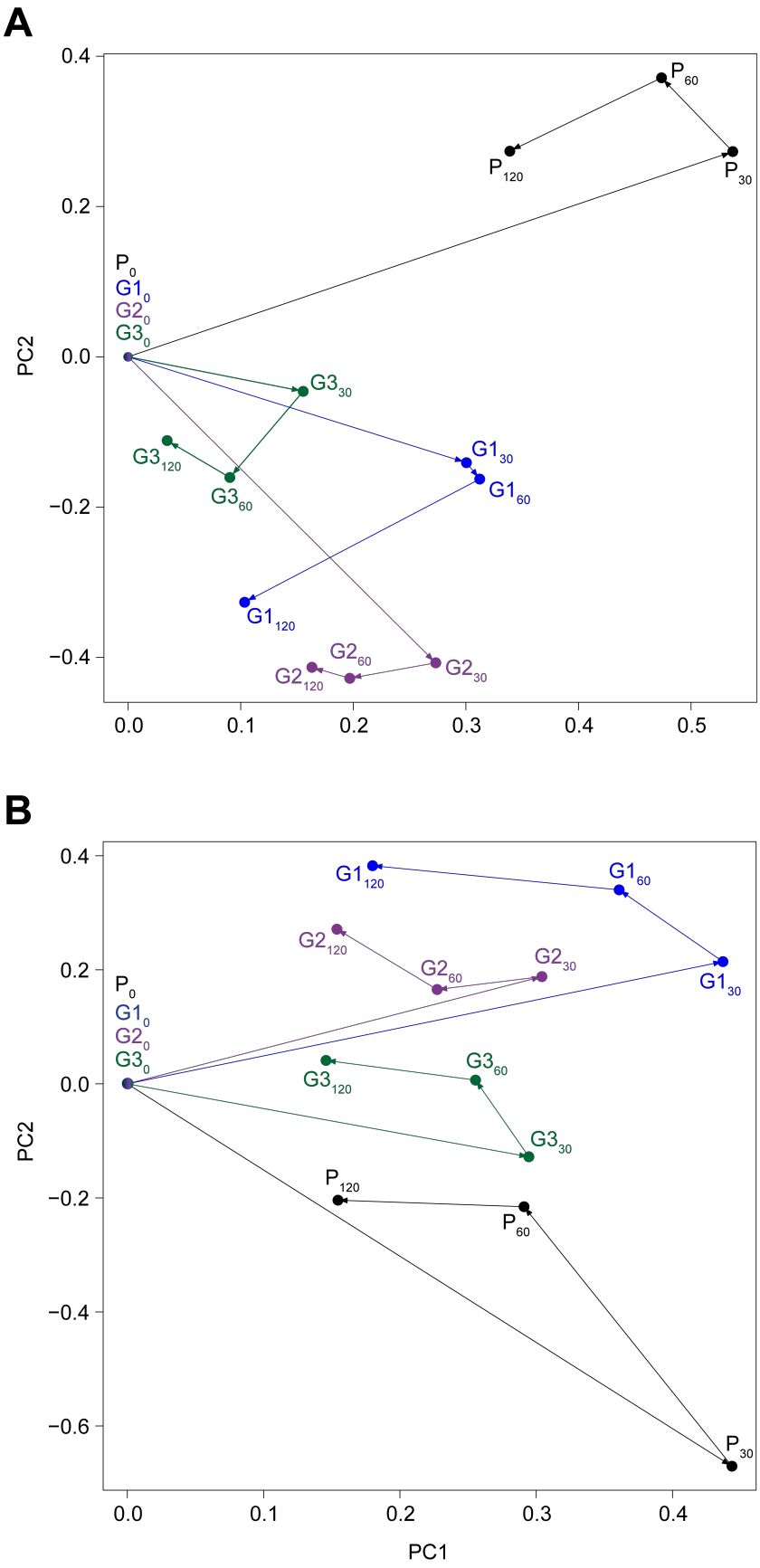


Figure 6.11: Multivariate analyses of the time-course and associate CTAM datasets highlight the separate EGF-induced phosphorylation kinetics between MCF7-G and -P cells. (A) PCA of the phosphopeptides demonstrating a statistically significant change in abundance in at least one condition (as shown in Fig 6.9). (B) PCA of the CTAM analysis of the time-course dataset (as shown in Fig 6.10).

data and CTAM network analysis. The data shown in Fig 6.11 also suggested that, as seen in Table 6.3, MCF7-G3 responded the least to EGF stimulation. Taken together, these analyses suggested that the EGF-induced signalling responses, and the routes through which the EGF-induced signals were propagated in the resistant cells, were different to the parental cells, whilst also being different to each of the other resistant cell-lines. These observations further supported the hypothesis that the network was rewired in these cells and the heterogeneity of this rewiring and phenotypes observed in Fig 6.3-6.8.

6.5.2 Degradation and recycling of EGFR in GDC-0941-resistant cells

Due to the differences observed in the EGF-induced changes in signalling between the parental and resistant cells, it was hypothesised that this may be as a result of impaired EGFR degradation and recycling (and the equilibrium that exists between these two processes). This is because the rate at which EGFR is retrieved from the membrane and is trafficked through endosomes determines its signalling kinetics. Recent data has outlined the importance of Akt in these processes, and as PI3K activates Akt in response to growth factor stimulation, it was not unreasonable to hypothesise that chronic inhibition of PI3K may have an effect on EGFR trafficking[291]. To test the degradation kinetics of EGFR in the resistant and parental cells, Western blots were performed for EGFR under the same conditions as the experiment presented and discussed in Figs 6.9-6.11 (Fig 6.12).

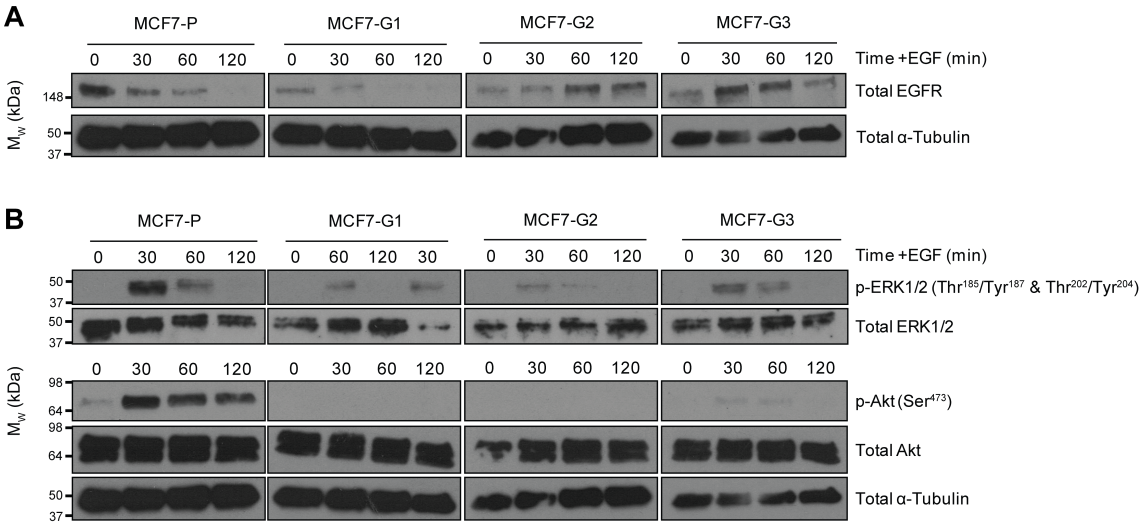


Figure 6.12: MCF7-G cells have impaired EGFR degradation and recycling kinetics and signal in an Akt-independent manner. Western blots for total EGFR (A) and activatory phosphorylation sites on ERK1/2 and Akt (B). Loading controls and blots for the total protein are shown in each case.

The data shown in Fig 6.12A demonstrated that the parental cells had a higher basal amount of EGFR under starvation conditions (t = 0 minutes), and upon stimulation with EGF, degraded EGFR to undetectable levels at 120 minutes. The resistant cells showed very different kinetics of EGFR degradation however: MCF7-G1 degrading EGFR more quickly, and MCF7-G2 and -G3 demonstrating an accumulation of EGFR. As observed in the MS data, the phosphorylation of ERK1/2 was stimulated in all four cell-lines and followed similar kinetics. As expected, and concordant with the data presented in Chapter 5, the phosphorylation of Akt was responsive to EGF stimulation in the parental cells. Consistent with the observations made in Fig 6.6, however, Akt phosphorylation (on Ser^{473}) was undetectable in each of the resistant cell-lines and did not respond to EGF stimulation. This suggested that the activation of downstream substrates of Akt-mTOR-PI3K seen in the CTAM network analysis (Fig 6.10) was independent of Akt phos-

phorylation and that the network had been rewired to facilitate phosphorylation of downstream substrates without the need for canonical Akt activation.

6.6 Conclusions

The experiments described in this Chapter aimed to utilise the previously defined kinase signalling network to investigate the changes in kinase signalling and the plasticity of this network in cells that had acquired resistance to the targeted inhibition of PI3K and mTORC1/2.

Firstly, it was determined whether the parental MCF7 cells (MCF7-P) were sensitive to the two compounds with which models of acquired resistance would be generated (GDC-0941 and KU-0063794). Cell viability (MTS) assay data for the parental cells treated with either GDC-0941 or KU-0063794 over a range of concentrations demonstrated that the parental cells were indeed sensitive to both compounds, although they were initially more sensitive to the former (Fig 6.2). Treating six identical populations of the parental MCF7 cells with an increasing concentration of each drug – three with GDC-0941 and three with KU-0063794 – led to the development of six cell-line models of acquired resistance (MCF7-G1:3 and MCF7-K1:3). Cell viability assays of the resistant cell-lines and the parental cells in the presence of 1 μM of the relevant compound revealed that these models were indeed resistant to the drugs with which they had been treated chronically (Fig 6.3). Consistent with the observation that the parental cells were less sensitive to KU-0063794 treatment, the cells resistant to this compound (MCF7-K1:3) were comparatively less so compared to the GDC-0941-resistant cells. These data also showed that there was a degree of heterogeneity in the relative resistance of each of the cell-lines resistant to the same compound.

The phosphoproteome of the six resistant cell-lines was then profiled by nLC-MS/MS and compared to that of the parental cells. To ensure the data were not confounded by acquired differences in the abundance of the unmodified proteins on which the measured phosphorylation sites were present, the total proteome of these cells was also compared and the phosphoproteomics data normalised to these differences. This comparison revealed that, on the whole, many statistically significant differences in the abundance of phosphopeptides and proteins between the resistant and parental cells existed (Tables 6.1 and 6.2). CTAM network analysis using these data showed that many branches of the defined network had altered activites in the resistant cells (Fig 6.4), and highlighted that the signalling pathways containing the targeted kinases (e.g. Akt-mTOR-PI3K and Akt-mTOR-p70S6K-PI3K) remained inhibited in the resistant cells. Accordant with other studies, this suggested that other signalling pathways were compensating for the lack of the activity of the inhibited kinases.

The patterns of the differences between the resistant and parental cells observed in the CTAM network analysis were distinct between the MCF7-G and -K cells. This was assessed objectively using multivariate analysis and this was able to separate the MCF7-G and -K cells both from the parental cells and from one another (Fig 6.5). This suggested that the evolved differences in the network's activity were distinct for cells treated with an mTORC1/2 or PI3K inhibitor. Moreover, this analysis showed a degree of separation between the three cell-lines resistant to same compound, in-line with the heterogeneity in proliferation rates observed in Fig 6.3, and suggestive of differences in the network's remodelling between cells treated with the same inhibitor.

The observation that the targeted kinases remained inhibited in these cells was validated by extracting some of their known substrates from the MS data and by Western blots for these (Fig 6.6). These experiments provided further evidence to support the observed maintained inhibition of PI3K and mTORC1/2 in the resistant cells, as the levels of Akt and p70S6K phosphorylation

were substantially reduced when compared to parental cells.

To investigate whether the changes in signalling network activity seen in the CTAM analysis had an impact of the phenotypes of the resistant cells, and to determine whether the heterogeneity observed in the remodelling of the network could be recapitulated based on these phenotypes, the resistant and parental cells' viability was assayed under treatment with a panel of small-molecule kinase inhibitors (Fig 6.7). These experiments revealed that the resistant cells had different responses to the drug combinations when compared to the parental cells; this being in-line with the idea that their signalling network had been remodelled. Multivariate analyses of these data, mirroring that seen in the CTAM data, were able to robustly separate the MCF7-G and -K cells from their parental counterparts and one another, whilst also separating individual cell-lines resistant to the same compound (Fig 6.8).

Having providing significant evidence to suggest that the signalling network was remodelled in the resistant cell-lines (Figs 6.4-6.8), it was next sought to determine whether this remodelling was merely a result of changes in the abundance of individual CTAM groups, or whether signals were flowing through different routes in the network. To achieve this, the phosphoproteomes of the parental and MCF7-G cells were assessed following treatment with EGF at four different time-points (Table 6.3). This analysis revealed widespread differences in the way in which EGF-induced signals propagated in these cells (Fig 6.9), as many phosphorylation sites demonstrated significantly different kinetics between the parental and resistant cells. PC analyses of these data and their corresponding CTAM network analysis led to clear separation of the resistant and parental cell time-points, indicating substantially different responses of the phosphoproteome to EGF stimulation and suggesting that signals were flowing through different routes in the network (Fig 6.10).

Finally, in order to provide a mechanistic basis for why the resistant cells responded to EGF differently, the degradation of EGFR was monitored over the EGF-stimulation time-course by Western blotting (Fig 6.11). This revealed that, although the parental cells degraded EGFR as would be expected[291], the resistant cells either degraded the receptor more quickly (MCF7-G1) or allowed its accumulation (MCF7-G2 and -G3). Western blots for a canonical readout of MAPK signalling activity showed that, as observed in the MS/CTAM data (Fig 6.10), this pathway responded to EGF treatment in the resistant and parental cells. Blots for a marker of Akt activity, however, showed that the activation of the Akt-mTOR-p70S6K-PI3K CTAM group seen in the MS data, was potentially independent of Akt activity. This further suggested that the resistant cells had rewired their signalling networks to overcome chronic inhibition of PI3K.

Taken together, these experiments showed that the defined kinase signalling network underwent significant remodelling upon acquisition of resistance to PI3K and mTORC1/2 inhibition. This remodelling was different for the two inhibitors, in agreement with their increasingly distinct functions of the kinases they target. Most strikingly, cells treated with the same inhibitor also evolved subtly different responses to the two drugs, indicating that there may be an element of stochasticity in the way in which signalling networks are rewired in response to targeted inhibitors.

Chapter 7

Discussion

PI3K and mTORC1/2, and the networks they signal within, manage cellular reponses to external and internal stimuli. Their roles are fundamental to the correct functioning of individual cells and the higher-order structures they form. These kinases are also relevant to cancer biology, as their dysregulation is closely linked to tumourigenesis. This project focused on three major areas of investigation in order to examine how chronic inhibition of these kinases influences signalling on a global level: (i) the development of sample preparation and mass spectrometric techniques for phosphoproteomics experiments; (ii) the empirical definition of a kinase signalling network and testing of its functional relevance; and (iii) how this network was remodelled in models of acquired resistance to PI3K or mTORC1/2-targeted therapies.

Mass spectrometry-based phosphoproteomics has found great utility in the study of protein phosphorylation; both in terms of localising modification sites on individual proteins, and studying their function(s). This is due, in part, to its high-content and unbiased nature, which allows investigative and discovery experiments that are not tractable with traditional molecular biology techniques. The success of phosphoproteomics experiments, however, is highly dependent on the upstream sample processing methods (i.e. enrichment of phosphopeptides from a more complex matrix) and the configuration of the instrumentation (e.g. MS^2 ion activation methods). Building on a previously described protocol, a more robust and efficacious phospho-enrichment method was developed (Chapter 3). In addition to this, several instrument-based conditions were optimised so as to produce the highest content phosphoproteomics datasets from the available instrumentation (Chapter 3).

The topology and plasticity of kinase signalling networks are integral to fundamental and disease biology. Consequently, the characterisation of these in diverse experimental settings is an important component of cell signalling research. Many network characterisation attempts have relied on multi-source protein and PTM interaction data from the literature, which, often limited, can result in inaccurate network models. In an effort to define a cell type-specific network, the phospho-enrichment and MS methods optimised in Chapter 3 were utilised to empirically define a kinase signalling network in MCF7 cells. This was achieved through the treatment of these cells with twenty structurally distinct kinase inhibitors and two phosphatase inhibitors, and their subsequent analysis by nLC-MS/MS (Chapter 4). The validity and functional relevance of this network and the underlying computational methods for its analysis were then confirmed through an analysis of the network's plasticity under well-characterised conditions (Chapter 5).

The way in which kinase signalling networks are remodelled in response to chronic, targeted kinase inhibition is a key aspect of acquired therapeutic resistance. Despite this, such remodelling has not been comprehensively investigated in an unbiased and global manner. Therefore, having established its functional relevance, the plasticity of the defined network was assessed in MCF7 cell-line models resistant to PI3K and mTORC1/2 inhibitors. This revealed widespread network remodelling that was not solely predicated on the initial conditions of the system and this raises interesting questions regarding the nature of the evolution of resistance (Chapter 6).

7.1 Phospho-enrichment and mass spectrometric methods for maximising phosphoproteome coverage

7.1.1 Principles of single-shot phospho-enrichment methods

As discussed in Chapter 1, phospho-enrichment methods are essential to the analysis of phosphorylated proteins and peptides from complex biological matrices. Many such methods – and the technologies they rely on – have undergone continuous development since their first applications. TiO_2 has been shown in numerous studies to be well-suited to the enrichment of phosphopeptides due to its relative ease of use, well understood retention mechanism, relatively high yield, and enrichment specificity [141,190,237,239,240,244,268].

"Single-shot" enrichments, where each sample is only enriched in a single dimension (e.g. not involving multi-dimensional liquid chromatographic fractionation), are essential to label-free phosphoproteomic workflows, as each sample is processed independently. Therefore, each single-extraction step must be optimised to afford the greatest depth of analysis possible and thus negate the lack of more in-depth fractionation of each sample. In addition, such methods should be highly reproducible and not lead to the introduction of significant technical bias. Single-shot enrichment methods have been investigated and, despite not reaching the depth afforded by 2D-LC methods, have proved efficacious for label-free approaches [141,244].

Due to the large numbers of samples that were to be compared in subsequent experiments, an effective single-shot enrichment method was crucial for this project. The protocol described in Montoya *et al.* [141] facilitated the identification of approximately 500 phosphopeptides per sample at ~90% specificity. Despite this, as discussed in Chapter 3, it was noted that a large amount of phosphopeptides enriched using this method were *N*-formylated – this being a potential source of quantitative and qualitative inaccuracies (Fig 3.1). In this study, this method was further tested to reduce the formation of this unwanted modification, whilst it was also sought to increase the number of identifications that could be made from a single extraction.

7.1.2 Technical considerations and limitations of the developed phospho-enrichment methods

7.1.2.1 Effects of on-line desalting on phosphopeptide recovery

In order to remove or reduce the presence of *N*-formylation on the recovered phosphopeptides, it was hypothesised that the C18 reversed-phase layer held within the micro spin-columns could be used to retain the phosphopeptides as they were eluted from the TiO_2. This would then allow their immediate desalting to remove residual NH_4OH, ammonium formate, and formic acid (Fig 3.2), and therefore significantly reduce the time the samples were exposed to formic acid.

The experiment performed in this study to compare the newly developed on-line method to that previously described (as in ref [141]) demonstrated that the former was able to drastically reduce the presence of *N*-formylated phosphopeptides, whilst increasing the number of identified phosphopeptides by approximately 1.6-fold. This increase in identifications resulted in a concomitant decrease in the enrichment specificity however, due to the co-purification of non-phosphorylated peptides (Fig 3.2). Inconsistent with the findings presented in Montoya *et al.*, an additional ammonium acetate wash did not improve the enrichment specificity (Fig 3.3). This was likely due to the fact that, although the ammonium acetate may have been effective in the removal of non-phosphorylated peptides from the TiO_2 layer, these may have then able to bind the C18 due to the relatively low proportion of ACN in the buffer (25%) and would therefore be present in the final

eluates.

7.1.2.2 Effects of off-line desalting on phosphopeptide recovery

The off-line method developed in this study was able to purify approximately 1.6-fold more phosphopeptides than the aforementioned on-line method, whilst simultaneously increasing the enrichment specificity (Fig 3.4). These data, corroborating other studies[292], also showed that GC (combined with C18) was more effective at retaining phosphopeptides than C18 alone. This is likely due to the additional ion-exchange retention mechanism afforded by GC aiding in the retention of hydrophilic moieties[292].

Further optimisation of this method – testing the effects of TiO_2 concentration, elution buffer hydrophobicity, and TiO_2 incubation buffers – demonstrated that increasing the ratio of TiO_2/protein and percentage of ACN in the elution buffer resulted in a greater number of phosphopeptide identifications, with no effect on the enrichment specificity (Fig 3.5A-B). The former finding, that a ratio of 50:1 TiO_2 (μg)/protein (μg) provided a substantial number of identifications without sacrificing specificity, is consistent with other studies[141]. Despite this, these data are at odds with other findings suggesting that a ratio of 2:1-8:1 TiO_2 (μg)/protein (μg) is optimal for phosphopeptide enrichment, as the higher the ratio, the lower the enrichment specificity[293]. The discrepancy between these datasets may in part be due to the presence of glycolic acid in the loading buffers used in the experiments shown in Fig 3.5 – this contributing to greater specificity of phosphopeptide binding to TiO_2 beads and therefore greater yield[239]. Taken together however, the data shown in Fig 3.5 and throughout Chapter 3 provide significant evidence to suggest that a ratio of 50:1 TiO_2 (μg)/protein (μg) yields a large number of phosphopeptides with high specificity, albeit with a slight increase in the material cost per enrichment.

The data in Fig 3.5 also show that, in agreement with published studies[239], the addition of 5% TFA in the TiO_2 loading/incubation buffer increased the specificity of enrichment (Fig 3.5C). A lack of TFA in this buffer, however, was compensated for by an increase in the concentration of the relevant α-hydroxy acid binding competitor (glycolic or lactic acid). This is contrary to data presented by Kettenbach and Gerber[240], which demonstrates that an increase in the proportion of TFA in the loading buffer reduced the number of identified phosphopeptides. These differences are likely explained by the fact that the experiments presented in Fig 3.5C were enrichments performed on total cell lysates as opposed to pre-SCX-fractionated samples enriched for populations of particularly acidic or basic phosphopeptides.

7.1.2.3 Effects of removing phosphopeptide acidification

The large majority of the current literature concerning TiO_2-MOAC-based phosphopeptide enrichment involves the acidification of phosphopeptides following their elution from the TiO_2 [141,244,241,239,247,294]. In order to determine whether this acidification step was unnecessary, and whether it was actually detrimental to the recovery of phosphopeptides, an acidification-free method ($\pm$ snap-freezing prior to vacuum-drying) was compared to the previously optimised off-line method (Fig 3.6). Consistent with the notion that the rate of phosphopeptide β-elimination (and therefore loss) is only significant under conditions of very high pH and high temperature[271] (such as those used for chemical derivatisation of phosphopeptides[221,222]), removal of the acidification and subsequent desalting steps from the enrichment protocol did not reduce the number of identified phosphopeptides (Fig 3.6). This observation only remained true for those samples that were snap-frozen immediately following NH_4OH-based elution from the TiO_2 beads, however. Unexpectedly, the removal of the acidification step did not increase the number of identified

phosphopeptides despite producing a visibly larger peptide pellet following vacuum-drying. The lack of difference in the number of phosphopeptides identified notwithstanding, it was clear that these two methods resulted in the enrichment of different portions of the phosphoproteome, as the samples produced using the newer method were much more complex in the hydrophobic region of their respective chromatograms (Fig 3.6D). The differences in the populations of phosphopeptides identified using these two methods were assessed further and this demonstrated that only 43% of identifications were common between the two methods; lower than can be accounted for by the stochastic nature of DDA-based experiments[274]. Moreover, these analyses showed that the method lacking acidification resulted in the enrichment of longer and more highly charged phosphopeptides (Fig 3.7).

The lack of increase in phosphopeptide identifications observed using the acidification-free method may have been a result of several factors. Firstly, as more highly charged phosphopeptides are less amenable to CID-based fragmentation methods[295], their increased recovery with the acidification-free method may have been to the detriment of the number that could be effectively sequenced by MS^2. Secondly, although a visibly larger peptide pellet was recovered with the newer method, these extra phosphopeptides may have been subsequently lost through the use of a low-organic (100% H_2O) and low pH (0.1% TFA) reconstitution buffer for MS analysis. Further work comparing different MS^2 ion activation methods, alongside different MS reconstitution buffers (e.g. a higher pH buffer such as triethyl ammonium bicarbonate with a higher proportion of organic solvent), may resolve the differences between the two enrichment schemes more clearly. Lastly, as the instrument may have been operating at capacity – i.e. the maximum number of ions that could be identified in a 45 min run with the utilised settings – a longer gradient length or more sensitive instrument may also resolve the differences between the two methods more effectively. Overall, however, the acidification-free method produced an equally substantial number of phosphopeptide identifications, whilst significantly reducing the protocol's time and material cost. Importantly, this rendered the comparison of large numbers of samples more amenable, thus lending itself to label-free experiments.

7.1.3 Principles of gas-phase fractionation

The selection of precursor ions in MS^1 scans is an essential process in the data-dependent acquisition of MS^2 data. This selection is predominantly based on the intensity and charge state of the precursor ions present in the initial survey scan. This creates a bias towards the most abundant precursor ions in each scan, therefore leading to the semi-random identification of only the most abundant proteins/peptides in a sample. Several alternative precursor ion selection strategies have been developed to overcome these limitations, including: DIA[163,164], exclusion and inclusion lists[272], and gas-phase fractionation[273]. Gas-phase fractionation relies on a restriction on the *m/z* of precursor ions that can be selected for MS^2. To this end, pre-selected *m/z* windows are used for each MS run, and precursor ions outside of these windows, however intense, are not be fragmented (Fig 3.8). Crucially for this project, as the MS^1 data used for label-free quantification are identical in each run, GPF could facilitate deeper analyses of the phosphoproteome without increasing overall MS analysis time.

7.1.3.1 Effectiveness and limitations of gas-phase fractionation

The effectiveness of gas-phase fractionation for label-free phosphoproteomics experiments was assessed and compared to full-scan precursor ion selection. The *m/z* windows were derived from

previous MS-based phosphoproteomics data produced from the same MCF7 cell-line. This experiment showed that, despite the instrument's ability to effectively fractionate precursor ions in the gas phase, this reduced the number of phosphopeptide identifications that could be made from an identical sample (Fig 3.8). This was unexpected as, theoretically, gas-phase fractionation should afford an increased depth of analysis, as lower abundance ions that may have been ignored in full-scan mode may then have a chance to be selected for fragmentation in each restricted *m/z* window. Several factors may explain this discrepancy between theory and observation. Firstly, the Orbitrap-XL instrument used for these experiments was operating with a single 100 ms micro-scan to produce each MS^2 spectra, corresponding to the fastest duty cycle available to the instrument. As noted by Scherl *et al.*, a lengthened duty cycle, utilising three 100 ms micro-scans, produces both higher quality and reproducible MS^2 spectra, whilst amplifying the gains afforded by GPF[273]. Secondly, as Fig 3.8F-G demonstrate, the instrument was not operating at full MS^2 occupancy, as many MS^1 scans produced only enough ions to warrant one or two subsequent fragmentation events (rather than the maximum of three). This may be due to software limitations of the instrument itself, or perhaps because the next most intense ions within the particular *m/z* window were not intense enough for their charge state to be determined and therefore were not selected. As a result, repetition of this experiment on a more sensitive, state-of-the-art instrument may provide more promising results. Finally, the number of ions identified in each gas phase fraction was heterogeneous, despite the rational design of the *m/z* windows from previously acquired data. Therefore, further optimisation of these windows based on these, or genomic data, may yield greater efficacy of the technique.

7.1.4 Principles of MS^2 ion activation methods

MS^2 fragmentation or ion activation methods are integral to the quality and quantity of the data output from LC-MS/MS experiments. Several independent groups have investigated various methods' utilities with regards to proteomics experiments[248–250,275]. These investigations have broadly shown that the optimal fragmentation technique is highly dependent on the peptide's physicochemical properties, as the PTMs present on the peptide and its charge state can heavily influence its fragmentation patterns. This has led to the development of hybrid fragmentation techniques such as EThcD[249,250] and decision-tree algorithms for choosing the most suitable method on-the-fly[251]. These methods are not yet widely available, however, and so the objective of this study was to compare the two most commonly employed methods for the fragmentation of phosphopeptides: CID-MSA and HCD.

7.1.4.1 Effectiveness and limitations of CID-MSA and HCD for phosphoproteomics analyses

The utility of CID-MSA and HCD methods was compared using an LTQ-Orbitrap-Velos instrument. To achieve this, four independent TiO_2 enrichments of an identical cell lysate were run in quadruplicate on the instrument. For each replicate injection a different MS^2 scheme was employed: (i) CID top 10, (ii) CID top 15), (iii) HCD top 7, and (iv) HCD top 10. It emerged from this comparison that, regardless of the top *N* spacing used, CID-MSA was able to identify a significantly greater number of unique phosphopeptides ions than HCD (Fig 3.9). This is consistent with other studies indicating that, due to the shortened duty cycle, CID-MSA provides substantially richer datasets on hydrid orbitrap instruments[248]. In addition, this comparison showed that, in-line with previously published data, the increase in identifications afforded by CID-MSA was accompanied by a reduction in the accuracy of phosphorylation site localisation and peptide identification[155,248] (Fig

3.9). This is most likely as a result of the higher mass accuracy and resolution of the fragment ion measurement (contributing to better charge state determination), wider measurable *m/z* range, and greater dynamic range afforded by HCD.

Contrary to the expected, this study also demonstrated that, independent of the ion activation method used, a greater number of MS^2 events between each MS^1 scan did not significantly increase the number of identifications made. A possible explanation for this could be that, as long (3 hour) gradients were utilised for the experiment, the instrument was fragmenting a large proportion of the available precursor ions even at lower top *N* spacings (e.g. CID-MSA top 10 and HCD top 7). Therefore, it could be hypothesised that repetition of this experiment with shorter LC gradients may provide greater resolution of the differences between an increase in top *N* spacing.

Overall, this study suggests – corroborating other datasets – that CID-MSA provides the greatest depth of phosphoproteome analysis on LTQ-Orbitrap-Velos instruments. The data also suggest that HCD produces more accurate peptide and phosphosite identifications. The increase in the number of phosphopeptide identifications facilitated through the use of CID-MSA (approximately 1.5-fold), however, outweighs the small increase in accuracy gained through the use of HCD when seeking to maximise the number of identifications that can be made from single extractions. In addition, due to the shorter duty cycle of CID-MSA methods, more MS^1 data-points are collected, thus allowing more accurate determination of precursor ion XIC peaks and therefore more accurate computation of their area and/or heights. Finally, as approximately 81% of the identifications made using HCD were common to those made using CID-MSA, the use of the latter does not exclude the analysis of a large proportion of the phosphoproteome (extracted with this phospho-enrichment method). This observation is seemingly only robust when analysing the most abundant phosphopeptide ions[155]; however, as the objective of this study was to optimise these methods for an unfractionated, single-shot extraction method, this was not as pertinent an issue.

7.1.5 Implications and future work

Taken together, the phospho-enrichment methods that were developed and presented in Chapter 3 facilitated the almost complete removal of unwanted *N*-formylation (Fig 3.2), whilst consistently increasing the number of phosphopeptide identifications when compared to the original method (Figs 3.3-3.6). This culminated in the ability to reliably identify approximately 1600 unique phosphopeptide ions from a single extraction derived from a 0.5 mg protein lysate. Considering that this only required a single, short MS run (45 minute gradient on an Orbitrap-XL instrument), this was competitive with other single-shot methods and even those combined with extensive sample fractionation[240,293]. The most important advantage of these methods however, was that, due to their simplicity and relative speed, they can be used in a relatively high-throughput manner to compare a large number of experimental conditions.

This study also reinforced the notion that CID-MSA fragmentation methods remain optimal for high-content phosphoproteomics experiments on early generation LTQ-Orbitrap hybrid instruments, despite the slight loss in peptide identification and PTM localisation accuracies. This is irrelevant on newer orbitrap instruments, however, as CID is either unavailable (e.g. Q-Exactive) or the optimal method can be selected on-the-fly (e.g. Fusion Tribrid instruments).

7.2 Empirical determination of a kinase signalling network

The topology of kinase signalling networks is critical to their function and output. Defining these topologies *de novo* is a significant analytical challenge and the majority of efforts to date have relied on the amalgamation of multiple literature sources[296,297]. Although informative, this approach is hampered by the fact that this amalgamation often obscures the cell- and context-specific connections that occur in true biological systems, therefore limiting the biological relevance of the resulting networks[191]. These 'averaged' or synthesised signalling networks, as often depicted in canonical signalling pathway diagrams, are derived from evolutionarily and functionally distant organisms and cell types. Therefore, these are not necessarily representative of biological kinase signalling networks. An advantage therefore of defining a signalling network empirically – using experimental data derived from a single system – is that cell-type and cell-context dependent connections are preserved; rendering the network a more reliable snapshot of kinase signalling and a valuable tool for the study of such a model.

Having established a highly specific and efficacious phospho-enrichment method amenable to the analysis of a relatively large sample numbers, alongside the comparison of mass spectrometric methods to allow maximal depth of analysis from single enrichments, the topology of a kinase signalling network was defined empirically in MCF7 cells. This was achieved through the measurement of the MCF7 phosphoproteome under treatment with a large array of structurally distinct kinase and phosphatase inhibitors targeting nodes of PI3K- and MEK-oriented signalling pathways (Table 2.2; Fig 4.1).

7.2.1 Technical considerations of the MS and computational methods

The network defining experiment consisted of the phosphoproteomic comparison of MCF7 cells treated with twenty-two structural distinct kinase/phosphatase inhibitors. This resulted in the quantitative comparison of 144 individual nLC-MS/MS runs. Combination of the identifications (as determined by Mascot) from these samples resulted in a database consisting of 13,405 unique phosphopeptide ions at an FDR = 5%. It is critical to control the peptide FDR at this stage, as false identifications can easily lead to misinterpretation of the biological meaning of such datasets. Many methods have been developed for the estimation of FDRs; however, the majority rely on the concomitant search of a randomised, shuffled, or reversed protein sequence database for the estimation of false positive identifications[298]. A quality control analysis of the identifications produced from this experiment demonstrated that, despite the threshold of 5%, the vast majority (>95%) of the identifications possessed a $q < 0.01$ (Fig 4.2A). Although such thresholds are largely arbitrary, and that protein identity would not be inferred from these peptides, a cut-off of 5% (corresponding to a $q < 0.05$) was deemed reasonable for this experiment.

MS^1-based label-free quantification strategies are highly dependent on liquid chromatographic stability and reproducibility. This is particularly true for the iteration of the PESCAL method used in this and other experiments, as sample-specific retention times for each phosphopeptide ion are calculated through the formation of linear models between the retention times of spiked-in enolase standard peptides. Therefore, although slight shifts in retention time of these can be easily tolerated, the elution order of these standard peptides must be consistent between nLC-MS/MS runs. An assessment of this and the overall effectiveness of the method was performed and this clearly demonstrated that: (i) the enolase peptide retention times aligned between individual runs in a reproducibly linear fashion (Fig 4.2B-C), and (ii) the elution order of these peptides was preserved across the 144 nLC-MS/MS runs (Fig 4.2E).

Despite consistent XIC peak height intensity distributions between the nLC-MS/MS runs, the quantitative data were quantile normalised to reduce the influence of technical variability and systematic biases on the dataset (Fig 4.2D). The process of quantile normalisation renders the statistical properties of each of the intensity distributions identical and has been shown to effectively reduce bias and variance whilst improving statistical power in similar datasets[265,299,300]. The magnitude and statistical significance of changes in phosphopeptide abundance across the 24 conditions (two biological replicates and three technical replicates for each) were assessed using a previously described linear modelling and empirical Bayesian shrinkage methods[262,266]. These analytical strategies have demonstrated significant utility in the analysis of microarray and proteomics datasets; both drastically reducing the number of false positive results and increasing statistical power in a practical and computationally efficient manner[266,283,284]. The results of this analysis demonstrated that the effect of small-molecule inhibitors of the targeted kinases on the phosphoproteome was heterogeneous (Table 4.1). This strongly supported the reasoning for the use of two inhibitors against each targeted kinase in the experiment, as this heterogeneity was observed even for inhibitors targeting the same kinase (e.g. Akt inhibitor VIII and MK-2206). As would be expected for inhibitors of kinase activity, each compound induced a greater number of significant reductions in phosphopeptide abundance than significant increases. Taken together, through the use of high quality quantitative data, combined with sophisticated statistical methods, robust changes in phosphopeptide abundance were identified from the raw MS data.

7.2.2 Validation of kinase inhibitor efficacy

The efficacy of targeted kinase inhibitors is often questionable, and some studies have demonstrated that incubation with an inhibitor can paradoxically increase the kinase's activity[301]. To provide evidence that the inhibitors employed in this study were indeed reducing the activity of their target kinase, Western blots were performed for their known downstream phosphorylation site substrates (where reliable reagents were available – Fig 4.3). These experiments showed that inhibitors of PI3K, mTORC1/2, PKC, EGFR, MEK and p70S6K each substantially reduced the phosphorylation of their respective substrates, with some showing complete abrogation below the limit of detection. Although blots for the total amount of unmodified substrate protein are not shown, the experiment was performed over a relatively short time-course (1 hour) and so these are unlikely to have changed significantly. Further work profiling the effects of the inhibitors not included in this analysis would provide further support to the MS data, although evidence of their efficacy has been described elsewhere[302–308].

7.2.3 Analysis of kinase relationships

In order to examine the similarities and differences between the effects of each of the compounds on the phosphoproteome, the phosphopeptides that demonstrated a statistically significant decrease in abundance (in at least one treatment) were extracted from the dataset. This filtered dataset was then deemed to be enriched for phosphorylation sites that were responsive to the inhibitory action of at least one of the surveyed compounds. The relationships between the effects of each of twenty-two compounds on this filtered dataset were then assessed by means of several multivariate analyses (Fig 4.4). These analyses suggested that inhibitors of the same kinase had very similar impacts on the filtered portion of the phosphoproteome as they correlated very strongly with each other and grouped closely in three-dimension PC-space. These data reinforced known relationships between the tested kinases, whilst the phosphatase (PP2A) inhibitors separated well from the other compounds in each analysis. For example, the effects of inhibitors of

mTOR, p70S6K and PI3K on the filtered phosphopeptides were strongly correlated and grouped well in three-dimensional PC-space (Fig 4.4A-C). This was also observed for inhibitors of the MAPK pathway (EGFR, MEK, and ERK).

Consistent with their canonical associations, inhibitors of Akt demonstrated statistically significant correlations with the inhibitors of mTOR and PI3K. Unexpectedly however, the Akt inhibitors partitioned more closely to CAMK2 and EGFR inhibitors in PC-space than to those of mTOR and PI3K. This was also observed in hierarchical clustering analysis of each inhibitor's quantitative effect on known consensus phosphorylation motifs (Fig 4.4D). Although somewhat counter-intuitive, this observation is consistent with several previous studies highlighting cross-talk between Akt/mTOR/PI3K and MAPK signalling in human malignancies[120,121], particularly via BRAF[122]. In addition, this phenomenon was unlikely to be as a result of experimental bias, as the linear modelling approach utilised to analyse the data included factors to account for this variability. Furthermore, as a total of six analytical replicates represented each data-point, this observation was statistically robust. Thus, although the confounding off-target effects of each inhibitor pair cannot be fully removed from these analyses, the high volume of replicates, alongside the inclusion of two inhibitors per kinase, aided in a substantial reduction of their overall impact.

7.2.4 Derivation of CTAMs from phosphoproteomics data

In this study, it was reasoned that the inclusion of two inhibitors per targeted kinase, alongside the robust statistical framework applied to the resulting data, would allow further classification of the phosphopeptides identified in the experiment. Consequently, it was hypothesised that phosphopeptides whose abundance was significantly reduced by both inhibitors targeting the same kinase would be enriched for markers of that kinase's activity. As the inhibitors possess off-target effects however, these populations of phosphopeptides would also represent the true kinases inhibited by each compound pair. Thus, the identified groups of phosphopeptides were dubbed compound-target activity markers (CTAMs).

A number of thresholds for CTAM identification were tested. As would be expected, an increase in the stringency of said thresholds reduced the number of CTAMs identified from the dataset (Fig 4.5). Due to the high-content nature of the experiment (afforded by the depth of analysis provided by the phospho-enrichment and mass spectrometric methods) it was reasoned that specificity was more critical than sensitivity. As a result, the most stringent thresholds were applied for the final analysis. The final thresholds controlled for false positive CTAM identifications on three levels: (i) phosphorylation site localisation probability (based on δ-score); (ii) magnitude of abundance change; and (iii) statistical robustness of abundance change. Most importantly, however, these criteria had to be satisfied for both inhibitors against the same kinase in order for a phosphopeptide to be identified as a CTAM (Fig 4.6).

Application of the CTAM classification algorithm to each inhibitor pair resulted in the identification of 610 CTAMs (Appendix 1). Analysis of these demonstrated that a large number of phosphopeptides were identified as being CTAMs of more than one kinase/inhibitor pair (Fig 4.7). Based upon these observations, the identified CTAMs were further classified into groups of CTAMs, according to which kinase activities they represented. For example, the phosphopeptide representing MAPK1 (Thr^{185} and Tyr^{187}) was determined to be a marker of both EGFR and MEK activities and was therefore classified into the EGFR-MEK CTAM group. Analysis of the data in this way produced 55 CTAM groups with varying numbers of members (Appendix 2). As would be expected, a number of known "signalling pathways" emerged from this analysis, including: Akt-mTOR-PI3K, Akt-mTOR-p70S6K-PI3K, and EGFR-MEK, and these served to validate

the approach. These groups contained several known substrates of these kinases (e.g. RICTOR Thr^{1135} in the Akt-mTOR-p70S6K-PI3K group[61]), alongside those previously undescribed. In addition to the known pathways represented in the data, a number of as yet uncharacterised connections between different kinases and inhibitors were also revealed (Fig 4.7).

Principal component analysis of the CTAM data enabled a more objective assessment of the relationships between the kinases interrogated in this study (Fig 4.7C). This demonstrated, similar to the analyses shown in Fig 4.4, that each inhibitor pair targeting the same kinase grouped well in PC-space, suggesting that their effects on the 610 CTAMs' quantitative behaviour were closely related. The unanticipated separation of Akt from PI3K and mTOR was also seen in these data; however, Akt, PI3K and mTOR possessed similar values in the third PC dimension (Fig 4.7C), suggesting a degree of similarity. Plotting the CTAM data in the form of a bipartite, undirected network graph further facilitated an assessment of how each of the kinases related to each other in these cells (Fig 4.8). Consistent with previous knowledge regarding their common signalling, and observations made within this study, this demonstrated that PI3K, mTOR, and p70S6K shared a large number of CTAMs. Similarly, MAPK-associated kinases also shared a large number of CTAMs with each other. Akt, as would be expected, shared a majority of its identified CTAMs with mTOR, PI3K, and p70S6K. In agreement with the observations noted in Figs 4.4, 4.6, and 4.7, Akt also possessed many common CTAMs with EGFR, CAMK2, and ERK. Most interestingly, a large number of Akt CTAMs were not shared with any other kinase node within the network, pointing towards a role for Akt outside of its canonical activators and other kinases probed in these experiments.

The role of Akt signalling outside of its most well-characterised partners PI3K and mTOR – and indeed *vice versa* – is a currently understudied aspect of these kinases' biology. Several studies, concordant with the observations made in Chapter 4, have hinted towards a greater role of PI3K/mTOR-independent Akt signalling and Akt-independent PI3K/mTOR signalling than has been previously appreciated[309]. For example, Akt has been shown to interact with, and be activated by, several tyrosine kinases in a PI3K-independent manner (e.g. Ack1/TNK2, PTK6, and Src). These kinases have been shown to phosphorylate Akt at sites within its kinase domain (e.g. Tyr^{176} and $Tyr^{315/326}$), independent of those canonically associated with its activity (i.e. Thr^{308} and Ser^{473})[310–312]. Thus, the data presented in Chapter 4 provide further evidence to suggest that Akt's signalling is not solely defined by its associations with PI3K and mTOR, nor does its association with MAPK signalling fully account for this.

7.2.5 CTAM identification: limitations of the methodology

The CTAM identification criteria discussed above were justified on the basis that a true marker of the activity of a kinase should demonstrate a modest reduction in its abundance under treatment with an inhibitor of said kinase. Consequently, a $\log_2$ fold-ratio $\leq$ -1 (adjusted $P \leq 0.1$), or a fold-ratio $\leq$ -0.75 with greater significance (adjusted $P \leq 0.05$) were deemed reasonable. Despite the success of this algorithm in identifying CTAMs in this experiment (Fig 4.6B), it could be argued that the magnitude of abundance change for each of the inhibited phosphopeptides is dependent on the time the cells were incubated with the relevant inhibitor. Indeed, the rate at which a phosphorylation site decreases in abundance with kinase inhibition (i.e. reducing k_{phos}; Fig 1.16) is also dependent on the rate at which that phosphorylation site is removed by phosphatases (i.e. k_{-phos}; Fig 1.16). Thus, the CTAMs identified within these experiments are limited by the fact that they are inherently biased towards phosphorylation sites which are relatively quickly removed by their phosphatase. As a result, variation in the incubation time used for each kinase inhibitor may

alter the CTAMs, and therefore network, which can be identified using this methodology.

It may also be argued that in order to improve the specificity of CTAM identification and reduce the impact of the off-target effects of each compound, a greater number of compounds or complementary RNAi technologies could be used[313]. A number of issues present themselves with this argument however. Firstly, although label-free MS methodologies allow the comparison of a theoretically unlimited number of samples, this is severely limited by the time-consuming manual processing of individual samples. Improvements in, and increasing availability of, robotic liquid handling systems will greatly improve this however[294]. Secondly, it remains the case that for a large number of kinases, the availability of structurally distinct and specific inhibitors is often poor. Finally, although an attractive alternative, RNAi technologies are potentially flawed for such signalling studies as they abolish the non-catalytic functions that many kinases provide (reviewed in ref[314]). Thus, as RNAi reduces the absolute expression of the targeted kinase (and often requires a long incubation time with the RNA to achieve this), this can cause more widespread perturbations, therefore complicating downstream analysis of the data.

7.2.6 Implications and future work

This study, whilst highlighting the utility of discovery-based MS methods for these types of experiments and the efficacy of label-free methods in this context, serves to expand our knowledge regarding the complexity of the relationships that exist between kinases and the networks they form. These data also illustrate that, despite some previous exploration of the cross-talk between these kinases and the networks the constitute[120–122], our knowledge of this complexity remains limited. The complexity of kinase signalling networks underpins their ability to receive, amplify, interpret, and respond to a large variety of biological signals[124]. A greater understanding of this will further our ability to manipulate these networks in the context of disease. Indeed, recent work concerning the combinatorial targeting of BRAF/MAPK and Akt/mTOR/PI3K signalling in certain cancers has proved fruitful, suggesting that this is an avenue worth pursuing[315,316]. As previously discussed, this study also strongly supports the notion that Akt plays a much greater role outside of its well-studied interactors than has been previously acknowledged. Further exploration of this may provide both deeper insight into Akt's functions, with the potential to expose novel therapeutic opportunities.

This study could be expanded through the characterisation of a greater number of kinases, as those profiled in these experiments only represent approximately 2% of those encoded by the human genome. Although limited by the availability and specificity of targeted kinase inhibitors, the system-specific CTAMs of many biologically important kinases (e.g. ALK, AMPK, PKA, GSK3β, Aurora A/B, etc.) could be profiled with this technology. Furthermore, due to the availability of isoform-specific inhibitors for some kinases, this study could also be expanded to include isoform-specific CTAMs (e.g. PI3Kα, β, and δ). As previously discussed, the increased availability of robotic technology, alongside the ever-increasing sensitivity and speed of mass spectrometers, will render larger experiments such as these more tractable in the future.

7.3 Analysis of signalling network plasticity in response to EGF/IGF-1 stimulation

Although the topology of kinase signalling networks underpins their basal function, signalling networks are not static structures. The dynamism of the strength of the connections that constitute the network – often termed plasticity – is key to the adaptability of biological systems in response

to intra- or extracellular perturbations[317] . The experiments presented in Chapter 5 aimed to demonstrate that the network defined in Chapter 4 was biologically relevant, and more importantly, whether it could be used to assess signalling plasticity in a global and unbiased manner. Crucially for this project, it was reasoned that if the aforementioned hypotheses were true, the network could be used to interrogate the changes in signalling that occur in response to chronic kinase inhibition and the subsequent development of acquired resistance.

7.3.1 Phosphoproteomics of growth factor-treated MCF7 cells

To test the aforementioned hypotheses, MCF7 cells were serum-starved and treated with EGF or IGF-1 over five time-points. The phosphoproteome of these cells was then measured by nLC-MS/MS and the CTAMs represented in the resulting data were extracted and analysed computationally in order to infer the activity of the signalling network.

Initial analyses of the quantitative mass spectrometry data revealed that EGF and IGF-1 had different effects on the phosphoproteome, as the number of significantly changing phosphopeptides at each time-point was heterogeneous between the two growth factors (Table 5.1). EGF and IGF-1 are known to stimulate the activity of MAPK signalling and a supervised extraction of known EGF and IGF-1-responsive phosphorylation sites representing this pathway demonstrated that the MS data were consistent with such an activation[286]. This was characterised by an acute increase in the abundance of phosphorylation sites on MAPK1/3 (ERK1/2), MAP2K2 (MEK2), RAF1, and SOS1. EGF and IGF-1 are also known to stimulate the activity of PI3K/Akt signalling; however, IGF-1 is thought to activate this more potently[287]. In-line with this idea, the MS data demonstrated a significant increase in the abundance of a phosphorylation site within the activation loop of Akt under both EGF and IGF-1 treatment; however, this increase was markedly more sustained under treatment with the latter (Fig 5.2). The observation that both MAPK and PI3K/Akt was stimulated in these cells, and the differences in Akt phosphorylation between the two growth factors, was validated with Western blots for these phosphorylation sites (Fig 5.3). Taken together, these data provided strong support for the validity of the MS data and subsequent analyses based on these.

7.3.2 Utility of CTAM-based network analysis of independent experiments

To demonstrate the utility of the defined signalling network in the unbiased assessment of signalling in the same system in an independent experiment, the temporal dynamics of each of the CTAM groups represented within the EGF and IGF-1-treated dataset were analysed systematically (Fig 5.4). This revealed that the identified CTAM groups (i.e. branches of the signalling network) were modulated by the two growth factors with the expected kinetics[286,287]. This served to validate the biological accuracy of the approach and the biological relevance of the defined network. Moreover, this demonstrated the ability of the technique to provide an informative snapshot of network activity for each condition derived from an independently generated dataset (Fig 5.5). Multivariate analysis of the CTAM network state at each time-point provided insight into the differences in the CTAM network's plasticity between EGF and IGF-1 treatment (Fig 5.6). This analysis, in-line with the observations made in Fig 5.2-5.5, showed that IGF-1 induced more sustained changes in the network's activity compared to EGF. As a whole, these analyses demonstrated that this method of network analysis was not only able to recapitulate the known impacts on PI3K/Akt/mTOR and MEK signalling, but was also able to resolve the differences in the kinetics of network plasticity induced by the two growth factors.

It can be argued that, if the CTAM groups representing the network are true reflections of biochemical signalling pathways, individual phosphorylation sites within each CTAM group should demonstrate consensus in their activation kinetics. Analysis of the intra-group variability in response to the two growth factors revealed that the majority of CTAMs showed relatively similar behaviour at each time-point (Fig 5.7). Importantly, visual assessment of both a densely and lightly populated CTAM group demonstrated that this observation was largely independent of group size (Fig 5.8).

7.3.3 Strengths and limitations of CTAM measurement

The monitoring of network activity through the use of empirically defined CTAMs has several advantages. This method of analysis confers the experimenter with the ability to assess global network dynamics in an untargeted manner and without bias towards particular branches of the network. The inference of signalling pathway activity from large-scale phosphoproteomics experiments is a significant technical challenge. Several bioinformatic approaches, based on literature-derived signalling networks have been used for this purpose (e.g. Ingenuity Pathway Analysis[318] and KEGG[297]). As previously discussed, however, these methods neglect cell-specific interactions between kinases and their substrates. Although efforts have been made to include context-specific interactions within these models[319], these still do not afford the resolution of changes at the level of individual phosphorylation sites. Therefore, as the CTAM-based profiling of the network is specific to the biological system under interrogation (i.e. MCF7-defined CTAMs are used to analyse independent experiments with MCF7 cells), and provides phosphorylation site-level resolution, this method of analysis provides a model closer to the biological reality. Furthermore, in contrast to other methods used for the untargeted profiling of kinase activites, such as multiplexed kinase inhibitor beads (MIBS)-MS – where the activity of individual kinases is probed and inferred in isolation[320,321] – the CTAM grouping method provides information regarding the circuitry of kinase activation.

Despite the advantages afforded by CTAM-based network analyses, several disadvantages present themselves. Firstly, although there may be a portion of CTAMs that are common between independent biological systems, to truly render the analysis context-specific, CTAMs must be empirically defined for each model to be studied. Although this is relatively straightforward in a well-controlled, well-established cell-line or easily cultured primary model, it would be challenging in an *in vivo* setting. Secondly, as the network currently only contains approximately 2% of theoretical human kinome, it is inherently biased towards a pre-determined selection of kinase activities. This, as previously discussed, would be improved with a more comprehensive analysis of kinases within the network using state-of-the-art instrumentation. Finally, as the magnitude and kinetics of the change in abundance of individual phosphorylation sites in response to external stimuli is tightly controlled – and indeed variations in signal intensity and frequency can have distinct downstream impacts[322] – computing the mean of these intensities in order to infer the activity of the group as a whole leads to a loss of this information. For example, the EGFR-MEK group (shown in Fig 5.8B) contained four quantifiable CTAMs in this experiment. Although the temporal kinetics of each of the four CTAMs followed an almost identical increase in abundance and subsequent decay (suggesting that they do indeed represent a biochemical pathway), the magnitude of these changes is substantially different. Despite the fact that these differences may be integral to the function of the pathway, this information is lost when the mean, and therefore total group activity, is computed (Fig 5.8B, red line).

7.3.4 Implications and future work

The kinase signalling network defined in Chapter 4 was utilised to successfully profile EGF and IGF-1-induced network plasticity in an untargeted manner. These data serve to validate both the method by which the network was defined, and provides proof-of-concept of measuring its plasticity in independent MS experiments. Therefore, this provides a platform by which this signalling network can be profiled in these (and potentially other) cell-line models, under any given condition.

Systems-level analyses of EGF-induced phosphorylation kinetics have been performed by several groups[286,323]; however, despite its importance in biology and disease, the global effects of IGF-1 on the phosphoproteome are less well-characterised. Thus, whilst validating the CTAM-based profiling of the signalling network, this study provides a valuable resource for the further interrogation of IGF-1's function.

In conjunction with the previously discussed expansion of the network with a greater number of kinases, the work in this study could be expanded upon through the further subsetting of EGF and IGF-1-responsive phosphorylation sites within each CTAM group. This would provide greater resolution of the effects of these growth factors, as although two phosphorylation sites may indeed belong to the same biological signalling pathway, they might not both be responsive to the same growth factor(s).

7.4 Plasticity of the defined kinase signalling network in models of acquired resistance

Targeted kinase inhibitors show great promise for the treatment of a number of malignancies. Such therapies, although not effective in 100% of patients due to intrinsic resistance, are often effective upon initial treatment. Unfortunately, initially responsive tumours almost inevitably acquire resistance to the relevant inhibitor and relapse as a consequence. The mechanisms of this form of drug resistance have been studied intensively in order to facilitate the rational design of drug alterations/combinations to prevent or reverse its occurrence. Despite the uncovering of several mechanisms by which cells can evade prolonged targeted kinase inhibition, it has become increasingly clear that the global remodelling and plasticity of signalling networks level plays a key role[97,191,320]. Thus, in this study, the plasticity of signalling networks in response to chronic, targeted kinase inhibition was examined through the use of the network defined and validated in Chapters 4 and 5.

7.4.1 Development of models of acquired resistance

In order to examine the effects of chronic PI3K and mTORC1/2 inhibition – by GDC-0941 and KU-0063794 respectively – models of acquired resistance to these two compounds were developed. To this end, a single parental MCF7 culture was homogenised and separated into six identical populations of cells. Three of these were then intially treated with 100 nM GDC-0941 and three with 100 nM KU-0063794 (Fig 6.1). A relatively low initial concentration of each drug was chosen so as to allow the majority of cells to survive the initial treatment (as this concentration was below the IC_{50} of both drugs – Fig 6.2) and to therefore minimise the outright selection of intrinsically resistant cells. The concentrations of each of the drugs was then gradually increased over each passage to a maximum of 1 μM. This maximum was selected so as to minimise confounding off-target effects of each inhibitor.

The resistance of the resulting cultures was determined by means of longitudinal proliferation,

MTS and CV cell viability assays (Fig 6.3). These assays demonstrated that each of the cultures had acquired resistance to the relevant inhibitor; however, a degree of heterogeneity was observed in the proliferation rates of each of the resistant cell-lines. In-line with the observation that the parental cells were initially more sensitive to GDC-0941, the cells treated with this inhibitor were comparatively more resistant than those treated with KU-0063794 (Fig 6.3C-D). Interestingly, alongside the validation of the cell-lines' acquired resistance to these compounds, these data demonstrated that, despite these cells' ability to proliferate and increase in absolute number in the presence of 1 μM drug (Fig 6.3C-D, purple bars), their metabolic capacity – as inferred by the MTS assay – was affected to a greater extent (Fig 6.3C-D, orange bars). This was particularly prominent with regards to the cells resistant to the mTORC1/2 inhibitor; this being consistent with mTOR's central role in metabolic regulation[49,54].

7.4.2 Phosphoproteomic, CTAM network and phenotypic analyses of the resistant cell-lines

Having established the drug resistance of the cell-line models, their phosphoproteomes and total proteomes were compared to that of the parental cells by mass spectrometry. This culminated in the reliable quantification of 11,307 unique phosphopeptide ions and 2,593 proteins across the seven cell-lines. The $\log_2$ fold-ratios (vs parental) for each phosphopeptide ion were subsequently normalised to the $\log_2$ fold-ratios (vs parental) for each protein (where these data were available), producing a stoichiometry ratio. Initial analyses of these data demonstrated that a large number of phosphopeptides and proteins were significantly altered in abundance in the resistant cells (Table 6.2). Consistent with published studies demonstrating remodelling of kinase signalling in response to chronic kinase inhibition[324,325], CTAM network profiling of these cells revealed widespread differences in network branch activity in comparison to the parental cells (Fig 6.4). Due to the depth of this analysis however, these data emphasise the extent to which signalling networks are globally remodelled in order to overcome the chronic inhibition of single nodes. These experiments also showed that, as substrates of the CTAM groups comprised of mTOR and PI3K (e.g. Akt-mTOR-PI3K and Akt-mTOR-p70S6K-PI3K) remained inhibited in the resistant cells, their acquired resistance was likely not as a result of reactivation of these kinases or an up-regulation of drug-efflux pump expression – both of which being common mechanisms of resistance to small-molecule inhibitors such as these – but rather via activation of compensatory branches of the network. These observations were supported through Western blot analysis of PI3K, Akt, and mTOR activation markers (Fig 6.6).

Further exploration of the CTAM network data (Fig 6.4) through the use of multivariate analyses revealed the extent of the differences in network activity between the resistant cells and the parental cells. This also highlighted the differences in network remodelling between the MCF7-G and -K cells (Fig 6.5). This strongly supported the idea that cells chronically treated with GDC-0941 remodelled their network differently to those treated with KU-0063794, despite PI3K and mTORC1/2's canonical association with one another. These data also suggested that, due to the separation of cell-lines resistant to the same drug in PC-space (e.g. MCF7-K1, -K2 and -K3), these cells had evolved subtly different responses to the same compound.

In order to examine whether remodelling of the defined signalling network in the resistant cells manifested itself in a phenotypic outcome, and whether the differences in network remodelling between the parental and resistant cells could be recapitulated with an independent technique, the responses in viability of each of the resistant cell-lines to a panel of small-molecule kinase inhibitors were tested. It was reasoned that, as the wiring of the signalling network defines the

sensitivity of cells to targeted kinase inhibition, remodelling of this network should alter the way in which the resistant cells responded to these drugs. These data demonstrated that, not only were the resistant cells' responses substantially different from those of the parental cells, but the differences between MCF7-G and -K cells, and those between cells resistant to the same drug, were maintained (Fig 6.7). Multivariate analyses of these data reinforced these observations; separating the parental cells, MCF7-G and MCF7-K cells from one another (Fig 6.8). These experiments lend further credence to the idea that cells treated chronically with GDC-0941 or KU-0063794 remodel their network differently, whilst promoting the notion that even cells treated with the same compound can evolve divergent responses. Taken together, the CTAM network and phenotypic profiling of the resistant cells accentuate the complex relationship between PI3K and mTORC1/2; one that is yet to be fully elucidated. These data are consistent however, with the observation that phosphorylation sites modulated by mTORC1/2 and PI3K inhibitors only partially overlapped (Chapter 4; Fig 4.8) and with recent data highlighting mTOR's independence from PI3K in some systems[326].

7.4.3 Rewiring of the signalling network in cells resistant to GDC-0941

The experiments presented in Section 6.4 provided evidence to suggest that the defined signalling network had been remodelled upon acquisition of kinase inhibitor resistance, and that this remodelling was distinct between parental, MCF7-G, and MCF-K cells. To determine whether the network was indeed rewired (i.e. signals being propagated via different routes) or whether this remodelling was as a result of a simple increase in abundance of particular CTAMs, EGF-induced changes in the phosphoproteome were compared between the parental and MCF7-G cell-lines (Table 6.3). Although these experiments were performed on a lower performance instrument (Orbitrap-XL), 4,228 unique phosphopeptide ions were able to be compared across the 16 conditions. Analysis of the phosphopeptides showing significant alterations in their abundance in conjunction with CTAM network profiling of these data, revealed that the temporal dynamics of EGF-responsive signals were substantially altered in the resistant cells when compared to the parental cell-line (Fig 6.9-6.10). Further examination of these using multivariate analyses corroborated these observations; separating the EGF-induced responses of the resistant cells from those of the parental cell-line (Fig 6.11). These analyses, mirroring the heterogeneity in network remodelling observed in Fig 6.5 and 6.8, also highlighted that the signalling networks of these cells relaxed back towards their basal states through profoundly different routes in PC-space. Although these phosphoproteomics data were not normalised to total protein amounts (and therefore changes in phosphopeptide abundance over time may have been as a result of changes in absolute protein amount), they still highlight that the EGF-induced signals were processed differently between the cell-lines. For example, even though it is not possible to determine whether a phosphorylation site increased in abundance after 120 minutes of EGF exposure is stoichiometrically valid, the fact that the protein's abundance may also be increasing (and not in the other cell-lines) demonstrates that there remains a difference in signal processing.

Having provided evidence to suggest that the signalling network had indeed been rewired in the MCF7-G cells and that this rewiring was different between these three cell-lines, it was hypothesised that this may have been as a result of a difference in EGFR internalisation and degradation. Consistent with recently published data, Western blots for total EGFR in the parental cells showed that the receptor is gradually degraded upon prolonged exposure to EGF[291] (Fig 6.12A). These data also showed however, that these kinetics were profoundly different in each of the MCF7-G cell-lines. Akt has been shown to control the trafficking and degradation of EGFR[291], and as

such, it is perhaps unsurprising that chronic inhibition of its upstream activator (PI3K) resulted in an interference of this process in MCF7-G2 and -G3 cells. Interestingly, these experiments suggested that MCF7-G1 cells were able to correctly degrade EGFR in a PI3K/Akt-independent manner, suggesting that the lack of Akt activity may have been compensated for by an alternate signalling pathway (summarised in Fig 7.1). Further work assessing the pathway responsible and investigation of the spatial processing of EGFR in these cells would elucidate this mechanism to a greater extent.

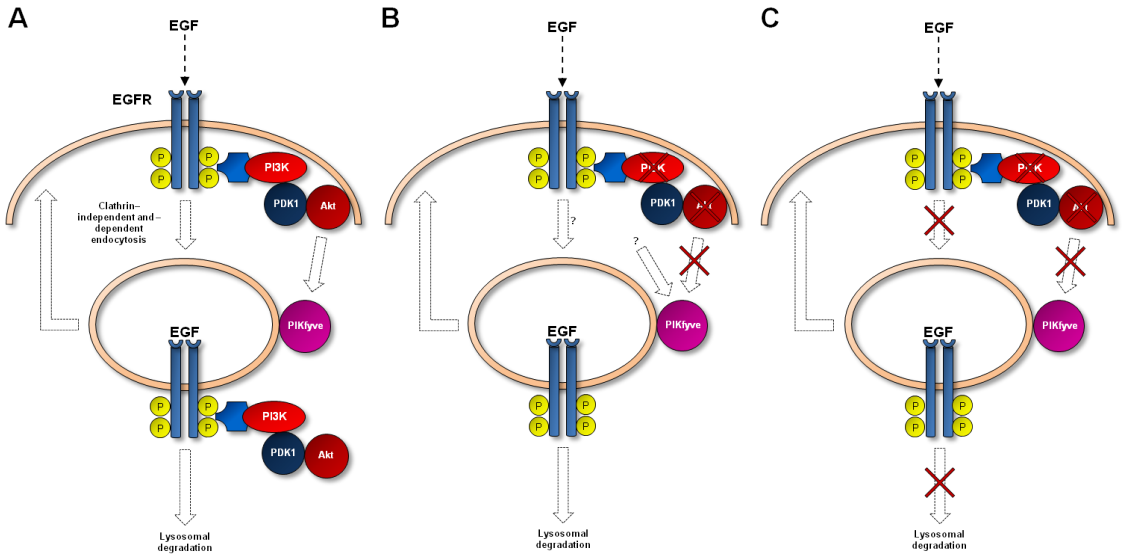


Figure 7.1: Models of EGFR recycling and degradation in parental and resistant cells. (A) Parental cells: EGFR dimerises and autophosphorylates upon binding to EGF. Receptor complexes are then trafficked into early endosomes where they continue to signal. Akt phosphorylation of PIKfyve aids in the correct endosomal progression of EGFR and its ultimate degradation. (B) MCF7-G1 cells: chronic inhibition of PI3K abrogates activation of Akt and therefore its phosphorylation of PIKfyve. The lack of Akt-dependent PIKfyve phosphorylation is potentially compensated for by unknown parallel processes and EGFR endosomal progression is restored. (C) MCF7-G2 and -G3 cells: chronic inhibition of PI3K abrogates activation of Akt and PIKfyve as in (B). EGFR does not properly progress through early and late endosomes and therefore is not degraded in the correct manner, leading to its accumulation.

Consistent with the mass spectrometry data summarised in Fig 6.9 and 6.10, ERK1/2 phosphorylation and activation remained responsive to EGF in the resistant cells (Fig 6.12B). In contrast, despite the clear activation of the Akt-mTOR-PI3K CTAM group observed in Fig 6.10, yet consistent with a maintained inhibition of PI3K (Fig 6.6B), phosphorylation of Akt at Ser^{473} was not responsive to EGF stimulation in these cells (Fig 6.12B). These data therefore suggest that the resistant cells had evolved a PI3K/Akt activation-independent manner by which the downstream substrates of these kinases could retain their responsiveness to growth factor stimulation.

7.4.4 Evolutionary routes to acquired resistance: lessons from chaos theory

Taken together, the experiments presented in Chapter 6 strongly supported the idea that the defined signalling network had been remodelled and rewired in the cells chronically treated with inhibitors of PI3K and mTORC1/2. These data suggested that, with respects to the GDC-0941-resistant cells, growth-stimulatory signals were re-routed through the network in order to maintain their proliferative effect(s) and bypass the continued inhibition of PI3K. These results are consistent with recent publications indicating that, rather than there being a single route by which cells can acquire resistance to a given therapy, there is a plethora of ways in which signalling networks can be rewired given prolonged kinase inhibition[327,328]. It is not known, however, whether the way in which cancer cells acquire resistance is predetermined by the molecular imprinting of cancer

cells at presentation – i.e. whether the acquisition resistance is entirely non-random (deterministic) or whether it involves some random elements (indeterministic). This understanding has potential implications for cancer therapeutics as, if the acquisition of resistance is deterministic and therefore reliant on genomic or other molecular factors present in tumour cells at presentation, then analysing the initial conditions of cancer cells could be used to predict the resistance mechanism likely to occur in such a tumour.

In this study, the same cell-line was split into six isogenic cultures, which were then maintained in the same concentrations of inhibitors by the same operator for the same amount of time and passage numbers. Despite this homogeneity in treatment, as previously discussed, the remodelling of the network and the subsequent sensitivity of the cells to multiple inhibitors was remarkably heterogeneous. Intra-tumoural heterogeneity has been documented at the genomic level[329] and this study suggests therefore, that diversity at the kinase signalling and functional levels also evolves as a result of chronic treatment with kinase inhibitors. Thus, as identical experimental conditions resulted in the evolution of distinctly remodelled signalling networks and drug-resistance phenotypes, the initial conditions of the system were not predictive of the evolutionary route that led to resistance. Although this study was performed in an *in vitro* system, it may be argued that the heterogeneity and complexity in *in vivo* systems can only be greater. Therefore, in support of an indeterministic model of acquired resistance, these data suggest that it may not be possible to predict the particular mechanism of resistance of a given tumour through knowledge of its molecular constitution at presentation. On the contrary, however, this apparent stochasticity may be as a result of deterministic chaos.

Chaos theory constitutes the study of the behaviour of highly dynamic and interconnected systems that are heavily dependent on their initial conditions – i.e. an arbitrarily-sized difference between the initial conditions of two otherwise identical systems can give rise to exponentially different future behaviours of each system (described by the Lyapunov exponent, Fig 7.2). Thus, although these systems may be fundamentally deterministic, the prediction of their future behaviour is exceptionally difficult. This phenomenon is therefore referred to as deterministic chaos. This has clear implications for the development of cancers (as has been previously investigated[330,331]); however, the application of models of this nature to the acquisition of resistance remains unexplored.

The data in Chapter 6 strongly suggest that, due to the complexity of the system and potentially stochastic variation of its evolution, the identification of resistance mechanisms at a global signalling level in individual tumours will require the ability to measure said network – with depth and without bias – *after* the emergence and detection of resistance. On the contrary, if the system is truly deterministic as stated by chaos theory, an increase of the accuracy with which the initial conditions can be determined may increase its predictability. In practice, however, due to the combinatorial complexity of the number of variables involved in tumourigenesis and resistance (e.g. genomic aberrations, epigenetic factors, kinase signalling dynamics, metabolic abnormalities and environmental factors), this may prove to be an impossible task. Therefore, the aforementioned notion of measuring these signalling networks by mass spectrometry once resistance has been acquired, may be a reasonable strategy in the management of this phenomenon. Further work determining the utility and power of this type of profiling in the prediction of rational secondary drug therapies, alongside advancements in mass spectrometric (and other) technologies, will determine whether this is indeed a viable approach. Alongside this, the application of chaotic mathematical models to the evolution of kinase signalling networks may provide a deeper understanding of how this occurs. Indeed, longitudinal analyses of these resistant cell-lines' or other resistance models' signalling network at different points in their evolution (i.e. Z_t), would provide

insight as to the predictability of these systems[332].

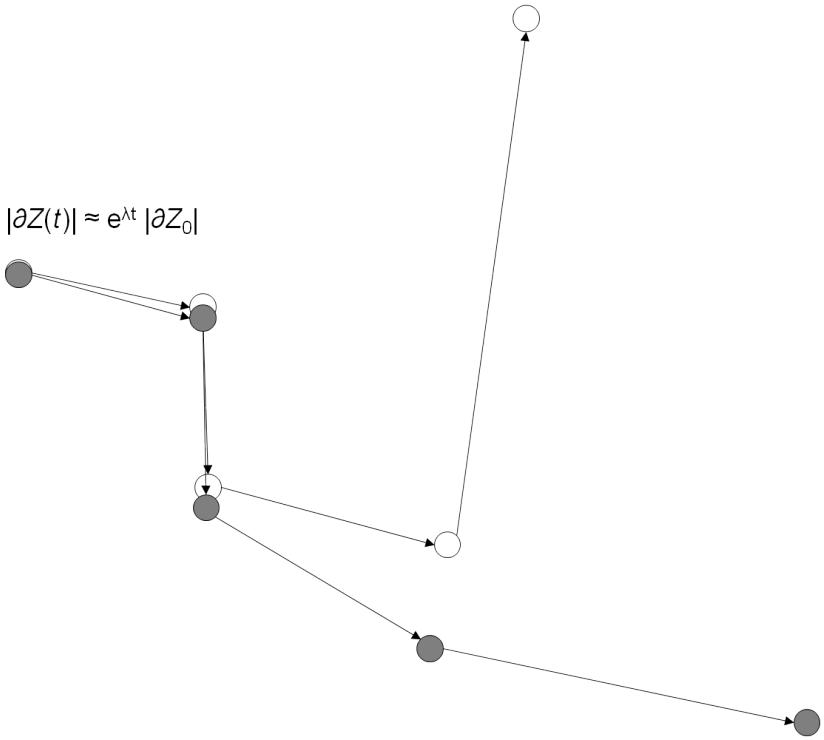


Figure 7.2: An illustration of deterministic chaos. Two seemingly identical systems (grey and white) start at defined points in space (Z_0) and possess a slight difference in their initial conditions (δZ_0). As the systems evolve (e.g. due to chronic treatment with a kinase inhibitor), they diverge as a function of time (t), at a rate defined by the Lyapunov exponent (λ)[333]. The rate of this change is dependent on the magnitude of δZ_0 and the initial spatial direction with which this divergence occurs. The inverse of this function, the Lyapunov time, represents the time for δZ_t to increase by a factor of e[332].

7.5 Concluding remarks

PI3K and mTORC1/2 are two kinases with great importance to human biology and disease. Therapeutic targeting of these and other kinases shows great promise in the continued improvement of cancer chemotherapeutics. Mass spectrometry-based phosphoproteomics has enabled the investigation of kinase signalling to depths that were previously impossible. The depth of such analyses has highlighted that kinase signalling is much more complex than has been previously appreciated and has demonstrated the utility of treating these systems as interconnected networks. Knowledge such as this continues to provide insights towards the ultimate goal of personalised cancer therapies. Furthermore, as the control signalling networks exert on biological processes is not restricted to cancer, these advances will have wide-reaching implications for other diseases.

The evolution of kinase signalling networks upon the acquisition of resistance to kinase inhibitor therapies has not yet been investigated at a global, network level. Therefore, the work within this thesis sought to first improve upon existing mass spectrometric and sample processing technologies for phosphoproteomics analyses, and to then apply these to the global study of acquired resistance.

The results of these investigations have provided insight into the architecture of kinase signalling networks and how these evolve in response to chronic single-node inhibition. These findings raise interesting questions about the nature of the acquisition of resistance and the implications of this for the future clinical management of this phenomenon. It remains to be determined whether these systems are truly deterministic and whether it is possible to predict resistance mechanisms at initial disease presentation. Although a challenging problem, the continued improvement of mass spectrometric technologies, alongside the potential for the application of chaotic mathematical models to these data, may provide a step towards answers to these questions.

References

[1] M. J. Zvelebil, L. MacDougall, S. Leevers, S. Volinia, B. Vanhaesebroeck, I. Gout, G. Panayotou, J. Domin, R. Stein, F. Pages, H. Koga, K. Salim, J. Linacre, P. Das, C. Panaretou, R. Wetzker, and M. Waterfield. Structural and functional diversity of phosphoinositide 3-kinases. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 351(1336): 217–223, 1996.

[2] B. Vanhaesebroeck, J. Guillermet-Guibert, M. Graupera, and B. Bilanges. The emerging mechanisms of isoform-specific PI3K signalling. *Nature Reviews Molecular Cell Biology*, 11 (5):329–341, 2010.

[3] B. Vanhaesebroeck, S. J. Leevers, K. Ahmadi, J. Timms, R. Katso, P. C. Driscoll, R. Woscholski, P. J. Parker, and M. D. Waterfield. Synthesis and function of 3-phosphorylated inositol lipids. *Annual Review of Biochemistry*, 70:535–602, 2001.

[4] L. C. Cantley. The phosphoinositide 3-kinase pathway. *Science*, 296(5573):1655–1657, 2002.

[5] M. J. Lindhurst, V. E. R. Parker, F. Payne, J. C. Sapp, S. Rudge, J. Harris, A. M. Witkowski, Q. Zhang, M. P. Groeneveld, C. E. Scott, A. Daly, S. M. Huson, L. L. Tosi, M. L. Cunningham, T. N. Darling, J. Geer, Z. Gucev, V. R. Sutton, C. Tziotzios, A. K. Dixon, T. Helliwell, S. O'Rahilly, D. B. Savage, M. J. O. Wakelam, I. Barroso, L. G. Biesecker, and R. K. Semple. Mosaic overgrowth with fibroadipose hyperplasia is caused by somatic activating mutations in PIK3CA. *Nature Genetics*, 44(8):928–933, 2012.

[6] Y. Samuels, Z. H. Wang, A. Bardelli, N. Silliman, J. Ptak, S. Szabo, H. Yan, A. Gazdar, D. M. Powell, G. J. Riggins, J. K. V. Willson, S. Markowitz, K. W. Kinzler, B. Vogelstein, and V. E. Velculescu. High frequency of mutations of the PIK3CA gene in human cancers. *Science*, 304(5670):554–554, 2004.

[7] A. Arcaro, M. J. Zvelebil, C. Wallasch, A. Ullrich, M. D. Waterfield, and J. Domin. Class II phosphoinositide 3-kinases are downstream targets of activated polypeptide growth factor receptors. *Molecular & Cellular Biology*, 20(11):3817–3830, 2000.

[8] J. Domin, L. Harper, D. Aubyn, M. Wheeler, O. Florey, D. Haskard, M. Yuan, and D. Zicha. The class II phosphoinositide 3-kinase PI3K-C2beta regulates cell migration by a PtdIns3P dependent mechanism. *Journal of Cell Physiology*, 205(3):452–462, 2005.

[9] I. Franco, F. Gulluni, C. C. Campa, C. Costa, J. P. Margaria, E. Ciraolo, M. Martini, D. Monteyne, E. De Luca, G. Germena, Y. Posor, T. Maffucci, S. Marengo, V. Haucke, M. Falasca, D. Perez-Morga, A. Boletta, G. R. Merlo, and E. Hirsch. PI3K class II alpha controls spatially restricted endosomal PtdIns3P and Rab11 activation to promote primary cilium function. *Developmental Cell*, 28(6):645–658, 2014.

[10] D. J. Gillooly, C. Raiborg, and H. Stenmark. Phosphatidylinositol 3-phosphate is found in microdomains of early endosomes. *Histochemistry and Cell Biology*, 120(6):455–453, 2003.

[11] R. Bago, N. Malik, M. J. Munson, A. Prescott, P. Davies, E. M. Sommer, N. Shpiro, R. Ward, D. Cross, I. G. Ganley, and D. R. Alessi. Characterisation of VPS34-IN1, a selective inhibitor of Vps34 reveals that the phosphatidylinositol 3-phosphate binding SGK3 protein kinase is a downstream target of Class III PI-3 kinase. *Biochemical Journal*, 463(3):413–427, 2014.

[12] J. M. Backer. The regulation and function of Class III PI3Ks: novel roles for Vps34. *Biochemical Journal*, 410(1):1–17, 2008.

[13] Y. Cao, Y. Wang, W. F. Abi Saab, F. Yang, J. E. Pessin, and J. M. Backer. NRBF2 regulates macroautophagy as a component of Vps34 Complex I. *Biochemical Journal*, 461(2):315–322, 2014.

[14] T. Nobukuni, M. Joaquin, o M. Rocci, S. G. Dann, S. Y. Kim, P. Gulati, M. P. Byfield, J. M. Backer, F. Natt, J. L. Bos, F. J. Zwartkruis, and G. Thomas. Amino acids mediate mTOR/raptor signaling through activation of class 3 phosphatidylinositol 3OH-kinase. *Proceedings of the National Academy of Sciences of the United States of America*, 102(40): 14238–14243, 2005.

[15] J. E. Slessareva, S. M. Routt, B. Temple, V. A. Bankaitis, and H. G. Dohlman. Activation of the phosphatidylinositol 3-kinase Vps34 by a G protein alpha subunit at the endosome. *Cell*, 126(1):191–203, 2006.

[16] J. Kim, Y. C. Kim, C. Fang, R. C. Russell, J. H. Kim, W. Fan, R. Liu, Q. Zhong, and K. L. Guan. Differential regulation of distinct Vps34 complexes by AMPK in nutrient stress and autophagy. *Cell*, 152(1-2):290–303, 2013.

[17] P. Voigt, C. Brock, B. Nurnberg, and M. Schaefer. Assigning functional domains within the p101 regulatory subunit of phosphoinositide 3-kinase gamma. *Journal of Biological Chemistry*, 280(6):5121–5127, 2005.

[18] J. Guillermet-Guibert, K. Bjorklof, A. Salpekar, C. Gonella, F. Ramadani, A. Bilancio, S. Meek, A. J. H. Smith, K. Okkenhaug, and B. Vanhaesebroeck. The p110 beta isoform of phosphoinositide 3-kinase signals downstream of G protein-coupled receptors and is functionally redundant with p110 gamma. *Proceedings of the National Academy of Sciences of the United States of America*, 105(24):8292–8297, 2008.

[19] P. Rodriguez-Viciana, C. Sabatier, and F. McCormick. Signaling specificity by Ras family GTPases is determined by the full spectrum of effectors they regulate. *Molecular & Cellular Biology*, 24(11):4943–4954, 2004.

[20] B. Geering, P. R. Cutillas, S. I. Nock, G.and Gharbi, and B. Vanhaesebroeck. Class IA phosphoinositide 3-kinases are obligate p85-p110 heterodimers. *Proceedings of the National Academy of Sciences of the United States of America*, 104(19):7809–7814, 2007.

[21] C-H. Huang, D. Mandelker, O. Schmidt-Kittler, Y. Samuels, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, S. B. Gabelli, and L. M. Amzel. The structure of a human p110 alpha/p85 alpha complex elucidates the effects of oncogenic PI3K alpha mutations. *Science*, 318 (5857):1744–1748, 2007.

[22] C. Jimenez, C. Hernandez, B. Pimental, and A. C. Carrera. The p85 regulatory subunit controls sequential activation of phosphoinositide 3-kinase by Tyr kinases and Ras. *Journal of Biological Chemistry*, 277(44):41556–41562, 2002.

[23] B. D. Cuevas, Y. Lu, M. Mao, J. Zhang, R. LaPushin, K. Siminovitch, and G. B. Mills. Tyrosine phosphorylation of p85 relieves its inhibitory activity on phosphatidylinositol 3-kinase. *Journal of Biological Chemistry*, 276(29):27455–27461, 2001.

[24] Y. H. Chiu, J. Y. Lee, and L. C. Cantley. BRD7, a Tumor Suppressor, Interacts with p85 alpha and Regulates PI3K Activity. *Molecular Cell*, 54(1):193–202, 2014.

[25] K. Ueki, D. A. Fruman, S. M. Brachmann, Y. H. Tseng, L. C. Cantley, and C. R. Kahn. Molecular balance between the regulatory and catalytic subunits of phosphoinositide 3-kinase regulates cell signaling and survival. *Molecular & Cellular Biology*, 22(3):965–977, 2002.

[26] K. Ueki, C. M. Fruman, D. A. Yballe, M. Fasshauer, J. Klein, T. Asano, L. C. Cantley, and C. R. Kahn. Positive and negative roles of p85 alpha and p85 beta regulatory subunits of phosphoinositide 3-kinase in insulin signaling. *Journal of Biological Chemistry*, 278(48): 48453–48466, 2002.

[27] L. Beltran, C. Chaussade, B. Vanhaesebroeck, and P. R. Cutillas. Calpain interacts with class IA phosphoinositide 3-kinases regulating their stability and signaling activity. *Proceedings of the National Academy of Sciences of the United States of America*, 108(39): 16217–16222, 2011.

[28] T. Maehama and J. E. Dixon. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *Journal of Biological Chemistry*, 273(22):13375–13378, 1998.

[29] E. A. Papakonstanti, A. J. Ridley, and B. Vanhaesebroeck. The p110 delta isoform of PI 3-kinase negatively controls RhoA and PTEN. *Embo Journal*, 26(13):3050–3061, 2007.

[30] W. E. Edimo, V. Janssens, E. Waelkens, and C. Erneux. Reversible Ser/Thr SHIP phosphorylation: a new paradigm in phosphoinositide signalling?: Targeting of SHIP1/2 phosphatases may be controlled by phosphorylation on Ser and Thr residues. *Bioessays*, 38(4): 634–642, 2012.

[31] C. G. Fedele, L. M. Ooms, M. Ho, J. Vieusseux, S. A. O'Toole, E. K. Millar, E. Lopez-Knowles, A. Sriratana, R. Gurung, G. G. Baglietto, L.and Giles, C. G. Bailey, J. E. J. Rasko, B. J. Shields, J. T. Price, P. W. Majerus, R. L. Sutherland, T. Tiganis, C. A. McLean, and C. A. Mitchell. Inositol polyphosphate 4-phosphatase II regulates PI3K/Akt signaling and is lost in human basal-like breast cancers. *Proceedings of the National Academy of Sciences of the United States of America*, 107(51):22231–22236, 2010.

[32] J. E. Harlan, P. J. Hajduk, H. S. Yoon, and S. W. Fesik. Pleckstrin homology domains bind to phosphatidylinositol-4,5-bisphosphate. *Nature*, 371(6493):168–170, 1994.

[33] S. J. Isakoff, T. Cardozo, J. Andreev, Z. Li, K. M. Ferguson, R. Abagyan, M. A. Lemmon, A. Aronheim, and E. Y. Skolnik. Identification and analysis of PH domain-containing targets of phosphatidylinositol 3-kinase using a novel in vivo assay in yeast. *Embo Journal*, 17(18): 5374–5387, 1998.

[34] G. Riddihough. More Meanders and Sandwiches. *Nature Structural Biology*, 1(11):755–757, 1994.

[35] C. S. Abrams, W. Zhao, and L. F. Brass. A site of interaction between pleckstrin's PH domains and G(beta gamma). *Biochimica Et Biophysica Acta-Molecular Cell Research*, 1314(3):233–238, 1996.

[36] M. Andjelkovic, D. R. Alessi, R. Meier, A. Fernandez, N. J. C. Lamb, M. Frech, P. Cron, P. Cohen, J. M. Lucocq, and B. A. Hemmings. Role of translocation in the activation and function of protein kinase B. *Journal of Biological Chemistry*, 272(50):31515–31524, 1997.

[37] H. Stocker, M. Andjelkovic, S. Oldham, M. Laffargue, M. P. Wymann, B. A. Hemmings, and E. Hafen. Living with lethal PIP3 levels: Viability of flies lacking PTEN restored by a PH domain mutation in Akt/PKB. *Science*, 295(5562):2088–2091, 2002.

[38] E. Patrucco, A. Notte, L. Barberis, G. Selvetella, A. Maffei, M. Brancaccio, S. Marengo, G. Russo, O. Azzolino, S. D. Rybalkin, L. Silengo, F. Altruda, R. Wetzker, M. P. Wymann, G. Lembo, and E. Hirsch. PI3K gamma modulates the cardiac response to chronic pressure overload by distinct kinase-dependent and -independent effects. *Cell*, 118(3):375–387, 2004.

[39] K. Lehmann, J. R. Mueller, B. Schlott, P. Skroblin, D. Barz, J. Norgauer, and R. Wetzker. PI3K gamma controls oxidative bursts in neutrophils via interactions with PKC alpha and p47phox. *Biochemical Journal*, 419:603–610, 2009.

[40] D. A. Guertin and D. M. Sabatini. Defining the role of mTOR in cancer. *Cancer Cell*, 12(1): 9–22, 2007.

[41] T. R. Peterson, M. Laplante, C. C. Thoreen, Y. Sancak, S. A. Kang, W. M. Kuehl, N. S. Gray, and D. M. Sabatini. DEPTOR is an mTOR inhibitor frequently overexpressed in multiple myeloma cells and required for their survival. *Cell*, 137(5):873–886, 2009.

[42] M. A. Frias, C. C. Thoreen, J. D. Jaffe, W. Schroder, T. Sculley, S. A. Carr, and D. M. Sabatini. mSin1 is necessary for Akt/PKB phosphorylation, and its isoforms define three distinct mTORC2s. *Current Biology*, 16(18):1865–1870, 2006.

[43] N. Hay and N. Sonenberg. Upstream and downstream of mTOR. *Genes & Development*, 18(16):1926–1945, 2004.

[44] A. C. Gingras, S. P. Gygi, B. Raught, R. D. Polakiewicz, R. T. Abraham, M. F. Hoekstra, R. Aebersold, and N. Sonenberg. Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes & Development*, 13(11):1422–1437, 1999.

[45] X. M. Ma and J. Blenis. Molecular mechanisms of mTOR-mediated translational control. *Nature Reviews Molecular Cell Biology*, 10(5):307–318, 2009.

[46] A. A. Michels, A. M. Robitaille, D. Buczynski-Ruchonnet, W. Hodroj, J. H. Reina, M. N. Hall, and N. Hernandez. mTORC1 directly phosphorylates and regulates human MAF1. *Molecular & Cellular Biology*, 30(15):3749–3757, 2010.

[47] J. Kim, M. Kundu, B. Viollet, and K. L. Guan. AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1. *Nature Cell Biology*, 13:132–141, 2011.

[48] X. Long, Y. Lin, S. Ortiz-Vega, K. Yonezawa, and J. Avruch. Rheb binds and regulates the mTOR kinase. *Current Biology*, 15(8):702–713, 2005.

[49] J. Avruch, K. Hara, Y. Lin, M. Liu, X. Long, S. Ortiz-Vega, and K. Yonezawa. Insulin and amino-acid regulation of mTOR signaling and kinase activity through the Rheb GTPase. *Oncogene*, 25(48):6361–6372, 2006.

[50] K. Inoki, Y. Li, T. Q. Zhu, J. Wu, and K. L. Guan. TSC2 is phosphorylated and inhibited by Akt and suppresses mTOR signalling. *Nature Cell Biology*, 4(9):648–657, 2002.

[51] E. B. Nascimento, M. Snel, B. Guigas, G. C. van der Zon, J. Kriek, J. A. Maassen, I. M. Jazet, M. Diamant, and D. M. Ouwens. Phosphorylation of PRAS40 on Thr246 by PKB/AKT facilitates efficient phosphorylation of Ser183 by mTORC1. *Cellular Signalling*, 22(6):961–967, 2010.

[52] E. V. Haar, S. Lee, S. Bandhakavi, T. Griffin, and D-H. Kim. Insulin signalling to mTOR mediated by the Akt/PKB substrate PRAS40. *Nature Cell Biology*, 9:316–323, 2007.

[53] R. C. Russell, Y. Tian, H. Yuan, H. W. Park, Y-Y. Chang, J. Kim, H. Kim, T. P. Neufeld, A. Dillin, and K-L. Guan. ULK1 induces autophagy by phosphorylating Beclin-1 and activating VPS34 lipid kinase. *Nature Cell Biology*, 15:741–750, 2013.

[54] D. M. Gwinn, D. B. Shackelford, D. F. Egan, M. M. Mihaylova, A. Mery, D. S. Vasquez, D. E. Turk, and R. J. Shaw. AMPK phosphorylation of raptor mediates a metabolic checkpoint. *Molecular Cell*, 30(2):214–226, 2008.

[55] X. Long, S. Ortiz-Vega, Y. S. Lin, and J. Avruch. Rheb binding to mammalian target of rapamycin (mTOR) is regulated by amino acid sufficiency. *Journal of Biological Chemistry*, 280(25):23433–23436, 2005.

[56] Y. Sancak, L. Bar-Peled, R. Zoucu, A. L. Markhard, S. Nada, and D. Sabatini. Ragulator-Rag Complex Targets mTORC1 to the Lysosomal Surface and Is Necessary for Its Activation by Amino Acids. *Cell*, 141(2):290–303, 2010.

[57] S. Katiyar, E. Liu, C. A. Knutzen, E. S. Lang, C. R. Lombardo, S. Sankar, J. I. Toth, M. D. Petroski, Z. Ronai, and G. G. Chiang. REDD1, an inhibitor of mTOR signalling, is regulated by the CUL4A–DDB1 ubiquitin ligase. *EMBO Reports*, 10(8):866–872, 2009.

[58] N. C. Wolff, S. Vega-Rubin-de Celis, X. J. Xie, D. H. Castrillon, W. Kabbani, and J. Brugarolas. Cell-type-dependent regulation of mTORC1 by REDD1 and the tumor suppressors TSC1/TSC2 and LKB1 in response to hypoxia. *Molecular & Cellular Biology*, 31(9):1870–1884, 2011.

[59] D. D. Sarbassov, D. A. Guertin, S. M. Ali, and D. M. Sabatini. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science*, 307(5712):1098–1101, 2005.

[60] J. M. Garcia-Martinez and D. R. Alessi. mTOR complex 2 (mTORC2) controls hydrophobic motif phosphorylation and activation of serum- and glucocorticoid-induced protein kinase 1 (SGK1). *Biochemical Journal*, 416(3):375–385, 2008.

[61] D. D. Sarbassov, S. M. Ali, D. H. Kim, D. A. Guertin, R. R. Latek, H. Erdjument-Bromage, P. Tempst, and D. M. Sabatini. Rictor, a novel binding partner of mTOR, defines a rapamycin-insensitive and raptor-independent pathway that regulates the cytoskeleton. *Current Biology*, 14(14):1296–1302, 2004.

[62] V. Zinzalla, D. Stracka, W. Oppliger, and M. N. Hall. Activation of mTORC2 by association with the ribosome. *Cell*, 144(5):757–768, 2011.

[63] J. Huang, S. Wu, C. L. Wu, and B. D. Manning. Signaling events downstream of mammalian target of rapamycin complex 2 are attenuated in cells and tumors deficient for the tuberous sclerosis complex tumor suppressors. *Cancer Research*, 69(15):6107–6114, 2009.

[64] P. Dalle Pezze, A. G. Sonntag, A. Thein, M. T. Prentzell, M. Godel, S. Fischer, E. Neumann-Haefelin, T. B. Huber, R. Baumeister, D. P. Shanley, and K. Thedieck. A dynamic network model of mTOR signaling reveals TSC-independent mTORC2 regulation. *Science Signaling*, 5(217):ra25, 2012.

[65] P. P. Hsu, S. A. Kang, J. Rameseder, Y. Zhang, K. A. Ottina, D. Lim, T. R. Peterson, Y. Choi, N. S. Gray, M. B. Yaffe, J. A. Marto, and D. M. Sabatini. The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. *Science*, 332(6035):1317–1322, 2011.

[66] P. Liu, W. Gan, H. Inuzuka, A. S. Lazorchak, D. Gao, D. Arojo, D. Liu, L. Wan, B. Zhai, Y. Yu, M. Yuan, B. Mo Kim, S. Shaik, S. Menon, S. P. Gygi, T. Ho Lee, J. M. Asara, B. D. Manning, J. Blenis, B. Su, and W. Wei. Sin1 phosphorylation impairs mTORC2 complex integrity and inhibits downstream Akt signalling to suppress tumorigenesis. *Nature Cell Biology*, 15: 1340–1350, 2013.

[67] D. Hanahan and R. A. Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5): 646–674, 2007.

[68] Y. Samuels, L. A. Diaz, O. Schmidt-Kittler, J. M. Cummins, L. DeLong, I. Cheong, C. Rago, D. L. Huso, C. Lengauer, K. W. Kinzler, B. Vogelstein, and V. E. Velculescu. Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell*, 7(6):561–573, 2005.

[69] I. G. Campbell, S. E. Russell, D. Y. H. Choong, K. G. Montgomery, M. L. Ciavarella, C. S. F. Hooi, B. E. Cristiano, R. B. Pearson, and W. A. Phillips. Mutation of the PIK3CA gene in ovarian and breast cancer. *Cancer Research*, 64(21):7678–7681, 2004.

[70] H. Zhang, G. Liu, M. Dziubinski, Z. Yang, S. P. Ethier, and G. Wu. Comprehensive analysis of oncogenic effects of PIK3CA mutations in human mammary epithelial cells. *Breast Cancer Research and Treatment*, 112(2):217–227, 2008.

[71] T. L. Yuan and L. C. Cantley. PI3K pathway alterations in cancer: variations on a theme. *Oncogene*, 27(41):5497–5510, 2008.

[72] B. Vanhaesebroeck, P. K. Vogt, and C. Rommel. PI3K: from the bench to the clinic and back. *Current topics in microbiology and immunology*, 347:1–19, 2010.

[73] L. Zhao and P. K. Vogt. Class I PI3K in oncogenic cellular transformation. *Oncogene*, 27 (41):5486–5496, 2008.

[74] Y. Mamane, E. Petroulakis, O. LeBacquer, and N. Sonenberg. MTOR, translation initiation and cancer. *Oncogene*, 25(48):6416–6422, 2006.

[75] K. Inoki, T. Q. Zhu, and K. L. Guan. TSC2 mediates cellular energy response to control cell growth and survival. *Cell*, 115(5):577–590, 2003.

[76] K. Tanaka, I. Babic, D. Nathanson, D. Akhavan, D. Guo, B. Gini, J. Dang, S. Zhu, H. Yang, J. De Jesus, A. N. Amzajerdi, Y. Zhang, C. C. Dibble, H. Dan, A. Rinkenbaugh, W. H. Yong, H. V. Vinters, J. F. Gera, W. K. Cavenee, T. F. Cloughesy, B. D. Manning, A. S. Baldwin, and P. S. Mischel. Oncogenic EGFR signaling activates an mTORC2-NF-kappa B pathway that promotes chemotherapy resistance. *Cancer Discovery*, 1(6):524–538, 2011.

[77] J. Zhang, P. L. Yang, and N. S. Gray. Targeting cancer with small molecule kinase inhibitors. *Nature Reviews Cancer*, 9(1):28–39, 2009.

[78] J. M. Garcia-Martinez, J. Moran, R. G. Clarke, A. Gray, S. C. Cosulich, C. M. Chresta, and D. R. Alessi. Ku-0063794 is a specific inhibitor of the mammalian target of rapamycin (mTOR). *Biochemical Journal*, 421(1):29–42, 2009.

[79] A. J. Folkes, K. Ahmadi, W. K. Alderton, S. Alix, S. J. Baker, G. Box, I. S. Chuckowree, P. A. Clarke, P. Depledge, S. A. Eccles, L. S. Friedman, A. Hayes, T. C. Hancox, A. Kugendradas, L. Lensun, P. Moore, A. G. Olivero, J. Pang, S. Patel, G. H. Pergl-Wilson, F. I. Raynaud, A. Robson, N. Saghir, L. Salphati, S. Sohal, M. H. Ultsch, M. Valenti, H. J. Wallweber, N. C. Wan, C. Wiesmann, P. Workman, A. Zhyvoloup, M. J. Zvelebil, and S. J. Shuttleworth. The identification of 2-(1H-indazol-4-yl)-6-(4-methanesulfonyl-piperazin-1-ylmethyl)-4-morpholin-4-yl-thieno[3,2-d]pyrimidine (GDC-0941) as a potent, selective, orally bioavailable inhibitor of class I PI3 kinase for the treatment of cancer. *Journal of Medicinal Chemistry*, 51(18):5522–5532, 2008.

[80] S. Alexander and P. Friedl. Cancer invasion and resistance: interconnected processes of disease progression and therapy failure. *Trends in Molecular Medicine*, 18(1):13–26, 2012.

[81] M. B. Meads, R. A. Gatenby, and W. S. Dalton. Environment-mediated drug resistance: a major contributor to minimal residual disease. *Nature Reviews Cancer*, 9(9):665–A674, 2009.

[82] J. Qi, M. A. McTigue, A. Rogers, E. Lifshits, J. G. Christensen, P. A. Jaenne, and J. A. Engelman. Multiple mutations and bypass mechanisms can contribute to development of acquired resistance to MET inhibitors. *Cancer Research*, 71(3):1081–1091, 2011.

[83] T. S. Mok, Y.-L. Wu, S. Thongprasert, C.-H. Yang, D.-T. Chu, N. Saijo, P. Sunpaweravong, B. Han, B. Margono, Y. Ichinose, Y. Nishiwaki, Y. Ohe, J.-J. Yang, B. Chewaskulyong, H. Jiang, E. L. Duffield, C. L. Watkins, A. A. Armour, and M. Fukuoka. Gefitinib or Carboplatin-Paclitaxel in Pulmonary Adenocarcinoma. *New England Journal of Medicine*, 361:947–957, 2009.

[84] P. B. Chapman, A. Hauschild, C. Robert, J. B. Haanen, P. Ascierto, J. Larkin, R. Dummer, C. Garbe, A. Testori, M. Maio, D. Hogg, P. Lorigan, C. Lebbe, T. Jouary, D. Scahdendorf, A. Ribas, S. J. O'Day, J. A. Sosman, J. M. Kirkwood, A. M. Eggermont, B. Dreno, K. Nolop, J. Lu, B. Nelson, J. Hou, R. J. Lee, K. T. Flaherty, G. A. McArthur, and BRIM-3 Study Group. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *New England Journal of Medicine*, 364(26):2507–2516, 2011.

[85] F. Janku, J. J. Wheeler, S. N. Westin, S. L. Moulder, A. Naing, A. M. Tsimberidou, S. Fu, G. S. Falchook, D. S. Hong, I. Garrido-Laguna, R. Luthra, J. J. Lee, K. H. Lu, and R. Kurzrock. PI3K/AKT/mTOR Inhibitors in Patients With Breast and Gynecologic Malignancies Harboring PIK3CA Mutations. *Journal of Clinical Oncology*, 30(8):777–782, 2012.

[86] R. Fritsche-Guenther, F. Witzel, A. Sieber, R. Herr, N. Schmidt, S. Braun, T. Brummer, C. Sers, and N. Bluthgen. Strong negative feedback from Erk to Raf confers robustness to MAPK signalling. *Molecular Systems Biology*, 7(489), 2011.

[87] K. E. O'Reilly, F. Rojo, Q. B. She, D. Solit, G. B. Mills, D. Smith, H. Lane, F. Hofmann, D. J. Hicklin, D. L. Ludwig, J. Baselga, and N. Rosen. mTOR inhibition induces upstream receptor tyrosine kinase signaling and activates Akt. *Cancer Research*, 66(3):1500–1508, 2006.

[88] A. Carracedo, L. Ma, F. Teruya-Feldstein, J.and Rojo, L. Salmena, A. Alimonti, A. Egia, A. T. Sasaki, G. Thomas, S. C. Kozma, A.a Papa, C. Nardella, L. C. Cantley, J. Baselga, and P. P. Pandolfi. Inhibition of mTORC1 leads to MAPK pathway activation through a PI3K-dependent feedback loop in human cancer. *Journal of Clinical Investigation*, 118(9): 3065–3074, 2008.

[89] A. B. Turke, Y. Song, C. Costa, R. Cook, C. L. Arteaga, J. M. Asara, and J. A. Engelman. MEK Inhibition Leads to PI3K/AKT Activation by Relieving a Negative Feedback on ERBB Receptors. *Cancer Research*, 72(13):3228–3237, 2012.

[90] A. Prahallad, C. Sun, S. Huang, F. Di Nicolantonio, R. Salazar, D. Zecchin, R. L. Beijersbergen, A. Bardelli, and R. Bernards. Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature*, 483(7387):100–103, 2012.

[91] B. S. Carver, C. Chapinski, J. Wongvipat, H. Hieronymus, Y. Chen, S. Chandarlapaty, V. K. Arora, C. Le, J. Koutcher, H. Scher, P. T. Scardino, N. Rosen, and C. L. Sawyers. Reciprocal Feedback Regulation of PI3K and Androgen Receptor Signaling in PTEN-Deficient Prostate Cancer. *Cancer Cell*, 19(5):575–586, 2011.

[92] A. Britschgi, R. Andraos, H. Brinkhaus, I. Klebba, V. Romanet, U. Mueller, M. Murakami, T. Radimerski, and M. Bentires-Alj. JAK2/STAT5 Inhibition Circumvents Resistance to PI3K/mTOR Blockade: A Rationale for Cotargeting These Pathways in Metastatic Breast Cancer. *Cancer Cell*, 22(6):796–811, 2012.

[93] J. T. Garrett, M. G. Olivares, C. Rinehart, N. D. Granja-Ingram, V. Sanchez, A. Chakrabarty, B. Dave, R. S. Cook, W. Pao, E. McKinely, H. C. Manning, J. Chang, and C. L. Arteaga. Transcriptional and posttranslational up-regulation of HER3 (ErbB3) compensates for inhibition of the HER2 tyrosine kinase. *Proceedings of the National Academy of Sciences of the United States of America*, 108(12):5021–5026, 2011.

[94] S. Guo and G. E. Sonenshein. Forkhead box transcription factor FOXO3a regulates estrogen receptor alpha expression and is repressed by the Her-2/neu/phosphatidylinositol 3-kinase/Akt signaling pathway. *Molecular & Cellular Biology*, 24(19):8681–8690, 2004.

[95] S. P. Tenbaum, P. Ordonez-Moran, I. Puig, I. Chicot, O. Arques, S. Landolfi, Y. Fernandez, J. R. Herance, J. D. Gispert, L. Mendizabal, S. Aguilar, S. Ramon y Cajal, S. Jr. Schwartz, A. Vivancos, E. Espin, S. Rojas, J. Baselga, J. Tabernero, A. Munoz, and H. G. Palmer. Beta-catenin confers resistance to PI3K and AKT inhibitors and subverts FOXO3a to promote metastasis in colon cancer. *Nature Medicine*, 18(6):892–901, 2012.

[96] K. H. Paraiso, Y. Xiang, V. W. Rebecca, E. V. Abel, Y. A. Chen, A. C. Munko, E. Wood, I. V. Fedorenko, V. K. Sondak, A. R. Anderson, A. Ribas, M. D. Palma, K. L. Nathanson, J. M. Koomen, J. L. Messina, and K. S. Smalley. PTEN loss confers BRAF inhibitor resistance

to melanoma cells through the suppression of BIM expression. *Cancer Research*, 71(7): 2750–2760, 2011.

[97] P. Casado, J-C. Rodríguez-Prados, S. C. Cosulich, S. Guichard, B. Vanhaesebroeck, S. Joel, and P. R. Cutillas. Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Science Signaling*, 6 (268):rs6, 2013.

[98] P. Casado, M. P. Alcolea, F. Iorio, J-C. Rodríguez-Prados, B. Vanhaesebroeck, J. Saez-Rodríguez, S. Joel, and P. R. Cutillas. Phosphoproteomics data classify hematological cancer cell lines according to tumor type and sensitivity to kinase inhibitors. *Genome Biology*, 14(4):R37, 2013.

[99] K. P. Ng, A. M. Hillmer, C. T. H. Chuah, W. C. Juan, T. K. Ko, A. S. M. Teo, P. N. Ariyaratne, N. Takahashi, K. Sawada, Y. Fei, S. Soh, W. H. Lee, J. W. J. Huang, J. C. Allen, X. Y. Woo, N. Nagarajan, V. Kumar, A. Thalamuthu, W. T. Poh, A. L. Ang, H. T. Mya, G. F. How, L. Y. Yang, L. P. Koh, B. Chowbay, C. T. Chang, V. S. Nadarajan, W. J. Chng, H. Than, L. C. Lim, Y. T. Goh, S. L. Zhang, D. Poh, P. Tan, J. E. Seet, M. K. Ang, N. M. Chau, Q. S. Ng, D. S. W. Tan, M. Soda, K. Isobe, M. M. Nothen, T. Y. Wong, A. Shahab, X. A. Ruan, V. Cacheux-Rataboul, W. K. Sung, E. H. Tan, Y. Yatabe, H. Mano, R. A. Soo, T. M. Chin, W. T. Lim, Y. J. Ruan, and S. T. Ong. A common BIM deletion polymorphism mediates intrinsic resistance and inferior responses to tyrosine kinase inhibitors in cancer. *Nature Medicine*, 18(4):521–528, 2012.

[100] M. E. Gorre, M. Mohammed, K. Ellwood, N. Hsu, R. Paquette, P. N. Rao, and C. L. Sawyers. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science*, 293(5531):876–880, 2001.

[101] N. P. Shah, J. M. Nicoll, B. Nagar, M. E. Gorre, R. L. Paquette, J. Kuriyan, and C. L. Sawyers. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell*, 2(2):117–125, 2002.

[102] L. Choi, Y, M. Soda, Y. Yamashita, T. Ueno, J. Takashima, T. Nakajima, Y. Yatabe, K. Takeuchi, T. Hamada, H. Haruta, Y. Ishikawa, H. Kimura, T. Mitsudomi, Y. Tanio, H. Mano, and ALK Lung Cancer Study Group. EML4-ALK Mutations in Lung Cancer That Confer Resistance to ALK Inhibitors. *New England Journal of Medicine*, 363(18):1734–1739, 2010.

[103] T. Sasaki, J. Koivunen, A. Ogino, M. Yanagita, S. Nikiforow, W. Zheng, C. Lathan, J. P. Marcoux, J. Du, K. Okuda, M. Capelletti, T. Shimamura, D. Ercan, Y. Stumpfova, M.and Xiao, S. Weremowicz, M. Butaney, S. Heon, K. Wilner, J. G. Christensen, M. J. Eck, K-K. Wong, N. Lindeman, N. S. Gray, S. J. Rodig, and P. A. Jaenne. A novel ALK secondary mutation and EGFR signaling cause resistance to ALK kinase inhibitors. *Cancer Research*, 71(18): 6051–6060, 2011.

[104] S. Kobayashi, T. J. Boggon, T. Dayaram, P. A. Janne, O. Kocher, M. Meyerson, B. E. Johnson, M. J. Eck, D. G. Tenen, and B. Halmos. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *New England Journal of Medicine*, 352(8):786–792, 2005.

[105] H. A. Yu, M. E. Arcila, N. Rekhtman, C. S. Sima, M. F. Zakowski, W. Pao, M. G. Kris, V. A. Miller, M. Ladanyi, and G. J. Riely. Analysis of Tumor Specimens at the Time of Acquired

Resistance to EGFR-TKI Therapy in 155 Patients with EGFR-Mutant Lung Cancers. *Clinical Cancer Research*, 19(8):2240–2247, 2013.

[106] C. Montagut, A. Dalmases, B. Bellosillo, M. Crespo, S. Pairet, M. Iglesias, M. Salido, M. Gallen, S. Marsters, S. P. Tsai, A. Minoche, S. Somasekar, S. Serrano, H. Himmelbauer, J. Bellmunt, A. Rovira, J. Settleman, F. Bosch, and J. Albanell. Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer. *Nature Medicine*, 18(2):221–223, 2010.

[107] N. Ilic, T. Utermark, H. R. Widlund, and T. M. Roberts. PI3K-targeted therapy can be evaded by gene amplification along the MYC-eukaryotic translation initiation factor 4E (eIF4E) axis. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (37), 2011.

[108] N. Wagle, C. Emery, M. F. Berger, M. J. Davis, A. Sawyer, P. Pochanard, S. M. Kehoe, C. M. Johannessen, L. E. MacConaill, W. C. Hahn, M. Meyerson, and L. A. Garraway. Dissecting Therapeutic Resistance to RAF Inhibition in Melanoma by Tumor Genomic Profiling. *Journal of Clinical Oncology*, 29(22):3085–3096, 2011.

[109] N. Wagle, E. M. Van Allen, D. J. Treacy, D. T. Frederick, Z. A. Cooper, A. Taylor-Weiner, M. Rosenberg, E. M. Goetz, R. J. Sullivan, D. N. Farlow, D. C. Friedrich, K. Anderka, D. Perrin, C. M. Johannessen, A. McKenna, K. Cibulskis, G. Kryukov, E. Hodis, D. P. Lawrence, S. Fisher, G. Getz, S. B. Gabriel, S. L. Carter, K. T. Flaherty, J. A. Wargo, and L. A. Garraway. Dissecting Therapeutic Resistance to RAF Inhibition in Melanoma by Tumor Genomic Profiling. *Cancer Discovery*, 4(1):61–68, 2011.

[110] R. Nazarian, H. Shi, Q. Wang, X. Kong, R. C. Koya, H. Lee, Z. Chen, M.-K. Lee, N. Attar, H. Sazegar, T. Chodon, S. F. Nelson, G. McArthur, J. A. Sosman, A. Ribas, and R. S. Lo. Melanomas acquire resistance to B-RAF (V600E) inhibition by RTK or N-RAS upregulation. *Nature*, 468(7326):973–977, 2010.

[111] C. Montagut, S. V. Sharma, T. Shioda, U. McDermott, M. Ulman, L. E. Ulkus, D. Dias-Santagata, H. Stubbs, D. Y. Lee, A. Singh, L. Drew, D. A. Haber, and J. Settleman. Elevated CRAF as a potential mechanism of acquired resistance to BRAF inhibition in melanoma. *Cancer Research*, 68(12):4853–4861, 2008.

[112] P. I. Poulikakos, Y. Persaud, M. Janakiraman, X. Kong, C. Ng, G. Moriceau, H. Shi, M. Atefi, B. Titz, M. T. Gabay, M. Salton, K. B. Dahlman, M. Tadi, J. A. Wargo, K. T. Flaherty, M. C. Kelley, T. Misteli, P. B. Chapman, J. A. Sosman, T. G. Graeber, A. Ribas, R. S. Lo, N. Rosen, and D. B. Solit. RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF (V600E). *Nature*, 480(7377):387–390, 2012.

[113] J. A. Engelman, K. Zejnullahu, T. Mitsudomi, Y. Song, C. Hyland, J. O. Park, N. Lindeman, C.-M. Gale, X. Zhao, J. Christensen, T. Kosaka, A. J. Holmes, A. M. Rogers, F. Cappuzzo, T. Mok, C. Lee, B. E. Johnson, L. C. Cantley, and P. A. Janne. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signalling. *Science*, 316(5827): 1039–1043, 2007.

[114] A. Jahangiri, M. De Lay, L. M. Miller, W. S. Carbonell, Y.-L. Hu, K. Lu, M. W. Tom, J. Paquette, T. A. Tokuyasu, S. Tsao, R. Marshall, A. Perry, K. M. Bjorgan, M. M. Chaumeil, S. M. Ronen, G. Bergers, and M. K. Aghi. Gene expression profile identifies tyrosine kinase c-Met

as a targetable mediator of antiangiogenic therapy resistance. *Clinical Cancer Research*, 19(7):1773–1783, 2013.

[115] K. Yonesaka, K. Zejnullahu, I. Okamoto, T. Satoh, F. Cappuzzo, J. Souglakos, D. Ercan, A. Rogers, M. Roncalli, M. Takeda, Y. Fujisaka, J. Philips, T. Shimizu, O. Maenishi, Y. Cho, J. Sun, A. Destro, K. Taira, K. Takeda, T. Okabe, J. Swanson, H. Itoh, M. Takada, E. Lifshits, K. Okuno, J. A. Engelman, R. A. Shivdasani, K. Nishio, M. Fukuoka, M. Varella-Garcia, K. Nakagawa, and P. A. Jaenne. Activation of ERBB2 signaling causes resistance to the EGFR-direct therapeutic antibody cetuximab. *Science Translational Medicine*, 3(99): 99ra86, 2011.

[116] M. R. Girotti, M. Pedersen, B. Sanchez-Laorden, A. Viros, S. Turajlic, D. Niculescu-Duvaz, A. Zambon, J. Sinclair, A. Hayes, M. Gore, P. Lorigan, C. Springer, J. Larkin, C. Jørgensen, and R. Marais. Inhibiting EGF receptor or SRC family kinase signaling overcomes BRAF inhibitor resistance in melanoma. *Cancer Discovery*, 3(2):158–167, 2013.

[117] J. Villanueva, A. Vultur, J. T. Lee, R. Somasundaram, M. Fukunaga-Kalabis, A. K. Cipolla, B. Wubbenhorst, X. W. Xu, P. A. Gimotty, D. Kee, A. E. Santiago-Walker, R. Letrero, K. D'Andrea, A. Pushparajan, J. E. Hayden, K. D. Brown, S. Laquerre, G. A. McArthur, J. A. Sosman, K. L. Nathanson, and M. Herlyn. Acquired resistance to BRAF inhibitors mediated by a RAF kinase switch in melanoma can be overcome by cotargeting MEK and IGF-1R/PI3K. *Cancer Cell*, 18(6):683–695, 2010.

[118] J. A. Engelman, L. Chen, X. Tan, K. Crosby, A. R. Guimaraes, R. Upadhyay, M. Maira, K. McNamara, S. A. Perera, Y. Song, L. R. Chirieac, R. Kaur, A. Lightbown, J. Simendinger, T. Li, R. F. Padera, C. Garcia-Echeverria, R. Weissleder, U. Mahmood, L. C. Cantley, and K.-K. Wong. Effective use of PI3K and MEK inhibitors to treat mutant Kras G12D and PIK3CA H1047R murine lung cancers. *Nature Medicine*, 14(12):1351–1356, 2008.

[119] K. Walsh, M. S. McKinney, C. Love, Q. Liu, A. Fan, A. Patel, J. Smith, A. Beaven, D. D. Jima, and S. S. Dave. PAK1 mediates resistance to PI3K inhibition in lymphomas. *Clinical Cancer Research*, 19(5):1106–1115, 2013.

[120] C. Wang, A. Cigliano, S. Delogu, J. Armbruster, F. Dombrowksi, M. Evert, X. Chen, and D. F. Calvisi. Functional crosstalk between AKT/mTOR and Ras/MAPK pathways in hepatocarcinogenesis: implications for the treatment of human liver cancer. *Cell Cycle*, 12(13): 1999–2010, 2013.

[121] E-R. Lee, J. Y. Kim, Y. J. Kang, J. Y. Ahn, J. H. Kim, B. W. Kim, H. Y. Choi, M. Y. Jeong, and S. G. Cho. Interplay between PI3K/Akt and MAPK signaling pathways in DNA-damaging drug-induced apoptosis. *Biochimica et Biophysica Acta*, 1763(9):958–968, 2006.

[122] M. Cheung, A. Sharma, S. V. Madhunapantula, and G. P. Robertson. Akt3 and mutant V600E B-Raf cooperate to promote early melanoma development. *Cancer Research*, 68 (9):3429–3439, 2008.

[123] P. K. Kreeger and D. A. Lauffenburger. Cancer systems biology: a network modeling perspective. *Carcinogenesis*, 31(1):2–8, 2010.

[124] J. S. Logue and D. K. Morrison. Complexity in the signaling network: insights from the use of targeted inhibitors in cancer therapy. *Genes & Development*, 26(7):641–650, 2012.

[125] P. Casado, B. Bilanges, V. Rajeeve, B. Vanhaesebroeck, and P. R. Cutillas. Environmental stress affects the activity of metabolic and growth factor signaling networks and induces autophagy markers in MCF7 breast cancer cells. *Molecular & Cellular Proteomics*, 13(3): 836–848, 2014.

[126] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, and B. Kuster. Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry*, 389(4): 1017–1031, 2007.

[127] B. T. Chait. Mass spectrometry: bottom-up or top-down? *Science*, 314(5796):65–66, 2006.

[128] F. Kjeldsen, A. M. B. Giessing, C. R. Ingrell, and O. N. Jensen. Peptide sequencing and charaterization of post-translational modifications by enchanced ion-charging and liquid chromatography electron-transfer dissociation tandem mass spectrometry. *Analytical Chemistry*, 79(24):9243–9252, 2007.

[129] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse. Electrospray ionisation for mass spectrometry of large biomolecules. *Science*, 246(4926):64–71, 1989.

[130] S. Nguyen and J. B. Fenn. Gas-phase ions of solute species from charged droplets of solutions. *Proceedings of the National Academy of Sciences of the United States of America*, 104(4):1111–1117, 2007.

[131] M. Dole, L. L. Mack, R. L. Hines, R. C. Mobley, L. D. Ferguson, and M. B. Alice. Molecular beams of macroions. *Journal of Chemical Physics*, 49(5):2240–2249, 1968.

[132] M. Gamero-Castano. Direct measurement of ion evaporation kinetics from electrified liquid surfaces. *Journal of Chemical Physics*, 113(2):815, 2000.

[133] M. Wilm and M. Mann. Analytical properties of nanoelectrospray ion source. *Analytical Chemistry*, 68(1):1–8, 1996.

[134] E. T. Gangl, M. M. Annan, N. Spooner, and P. Vouros. Reduction of signal suppression effects in ESI-MS using a nanosplitting device. *Analytical Chemistry*, 73(23):55635–5644, 2001.

[135] M. Karas, D. Bachmann, U. Bahr, and F. Hillenkamp. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *International Journal of Mass Spectrometry and Ion Processes*, 78:53–68, 1987.

[136] S-E. Ong and M. Mann. Mass spectrometry-based proteomics turns quantitative. *Nature Chemical Biology*, 1(5):252–262, 2005.

[137] B. A. Mamyrin, V. I. Karataev, D. V. Shmikk, and V. A. Zagulin. The mass-reflectron, a new non-mnagnetic time-of-flight mass spectrometer with high resolution. *Journal of Experimental and Theoretical Physics*, 64(1):82, 1973.

[138] A. Makarov. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Analytical Chemistry*, 72(6):1156–1162, 2000.

[139] A. Moradian, A. Kalli, M. J. Sweredoski, and S. Hess. The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics*, 14(4-5):489–497, 2014.

[140] J. R. Wisniewski, A. Zougman, N. Nagaraj, and M. Mann. Universal sample preparation method for proteome analysis. *Nature Methods*, 6:359–362, 2014.

[141] A. Montoya, L. Beltran, P. Casado, J-C. Rodríguez-Prados, and P. R. Cutillas. Characterisation of a TiO_2 enrichment method for label-free quantitative phosphoproteomics. *Methods*, 54(4):370–378, 2011.

[142] B. Canas, C. Pineiro, E. Calvo, D. Lopez-Ferrer, and J. M. Gallardo. Trends in sample preparation for classical and second generation proteomics. *Journal of Chromatography A*, 1153(1-2):235–258, 2007.

[143] T. Glatter, C. Ludwig, E. Ahrne, R. Aebersold, A. J. Heck, and A. Schmidt. Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *Journal of Proteome Research*, 11(11):5145–5156, 2012.

[144] C. Wu, J. C. Tran, L. Zamdborg, K. R. Durbin, M. Li, D. R. Ahlf, B. P. Early, P. M. Thomas, J. V. Sweedler, and N. L. Kelleher. A protease for 'middle-down' proteomics. *Nature Methods*, 9: 822–824, 2012.

[145] M. Peng, N. Taouatas, S. Cappadona, B. van Breukelen, S. Mohammed, A. Scholten, and A. J. Heck. Protease bias in absolute protein quantitation. *Nature Methods*, 9(6):524–525, 2012.

[146] P. A. Faull, H. V. Florance, C. Q. Schmidt, N. Tomcyzk, P. N. Barlow, T. R. Hupp, P. V. Nikolova, and P. E. Barran. Utilising ion mobility-mass spectrometry to interrogate macromolecules: Factor H complement control protein modules 10–15 and 19–20 and the DNA-binding core domain of tumour suppressor p53. *International Journal of Mass Spectrometry*, 298(1):99–110, 2010.

[147] C. A. Scarff, J. R. Snelling, M. M. Knust, C. L. Wilkins, and J. H. Scrivens. New structural insights into mechanically interlocked polymers revealed by ion mobility mass spectrometry. *Journal of the American Chemical Society*, 134(22):9193–9198, 2012.

[148] U. H. Mistarz, J. M. Brown, K. F. Haselmann, and K. D. Rand. A simple setup for gas-phase H/D exchange mass spectrometry coupled to electron transfer dissociation and ion mobility for analysis of polypeptide structure on a liquid chromatographic timescale. *Analytical Chemistry*, 86(23):11868–11876, 2014.

[149] L. Earley, L. C. Anderson, D. L. Bai, C. Mullen, J. E. Syke, A. M. English, J. J. Dunyach, Stafford G. C. Jr., J. Shabanowitz, D. F. Hunt, and P. D. Compton. Front-end electron transfer dissociation: a new ionization source. *Analytical Chemistry*, 85(17):8385–8390, 2013.

[150] F. Marino, A. Cristobal, N. A. Binai, N. Bache, A. J. Heck, and S. Mohammed. Characterization and usage of the EASY-spray technology as part of an online 2D SCX-RP ultra-high pressure system. *Analyst*, pages –, 2014.

[151] R. K. Boyd. Linked-scan techniques for MS/MS using tandem-in-space instruments. *Mass Spectrometry Reviews*, 13(5-6):359–410, 1994.

[152] D. Despeyroux, J. Bordas-Nagy, and K. R. Jennings. Determination of the amino acid sequence of cystine-containing peptides by tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 5(4):156–159, 1991.

[153] J. M. Wells and S. A. McLuckey. Collision-induced dissociation (CID) of peptides and proteins. *Methods in Enzymology*, 402:148–185, 2005.

[154] P. Roepstorff and J. Fohlman. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry*, 11(11):601, 1984.

[155] N. Nagaraj, R. C. D'Souza, J. Cox, J. V. Olsen, and M. Mann. Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *Journal of Proteome Research*, 9(12):9786–9794, 2010.

[156] B. Paizs and S. Suhai. Fragmentation pathways of protonated peptides. *Mass Spectrometry Reviews*, 24(4):508–548, 2005.

[157] J. Zhang, L. Xin, B. Shan, W. Chen, M. Xie, D. Yuen, W. Zhang, Z. Zhang, G. A. Lajoie, and B. Ma. PEAKS DB: de novo sequencing assited database search for sensitive and accurate peptide information. *Molecular & Cellular Proteomics*, 11(4):M111.010587, 2011.

[158] D. N. Perkins, D. J. Pappin, D. M. Creasy, and J. S. Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18):3551–3567, 1999.

[159] T. Koenig, B. H. Menze, M. Kirchner, F. Monigatti, K. C. Parker, T. Patterson, J. J. Steen, F. A. Hamprecht, and H. Steen. Robust prediction of the MASCOT score from an improved quality assessment in mass spectrometric proteomics. *Journal of Proteome Research*, 7 (9):3708–3717, 2008.

[160] H. Choi and A. I. Nesvizhskii. False discovery rates and related statistical concepts in mass spectrometry-based proteomics. *Journal of Proteome Research*, 7(1):47–50, 2008.

[161] K. Jeong, S. Kim, and N. Bandeira. False discovery rates in spectral identification. *BMC Bioinformatics*, 13(Suppl 16):S2, 2012.

[162] A. Michalski, J. Cox, and M. Mann. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *Journal of Proteome Research*, 10(4):1785–1793, 2011.

[163] J. C. Silva, M. V. Gorenstein, G-Z. Li, J. P. C. Vissers, and S. J. Geromanos. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Molecular & Cellular Proteomics*, 5(1):144–156, 2006.

[164] L. C. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, R. Bonner, and R. Aebersold. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics*, 11(6):O111.016717, 2012.

[165] M. Bauer, E. Ahrne, A. P. Baron, T. Glatter, L. L. Fava, A. Santamaria, E. A. Nigg, and A. Schmidt. Evaluation of data-dependent and data-independent mass spectrometric workflows for sensitive quantification of proteins and phosphorylation sites. *Journal of Proteome Research*, epub(epub):epub, 2014.

[166] J. D. Egertson, A. Kuehn, G. E. Merrihew, N. W. Bateman, B. X. MacLean, Y. S. Ting, J. D. Canterbury, D. M. Marsh, M. Kellmann, V. Zabrouskov, Wu. C. C., and M. J. MacCoss. Multiplexed MS/MS for improved data-independent acquisition. *Nature Methods*, 10(8): 744–746, 2014.

[167] D. Broudy, T. Killeen, M. Choi, N. Shulman, D. R. Mani, S. E. Abbatiello, D. Mani, R. Ahmad, A. K. Sahu, B. Schilling, K. Tamura, Y. Boss, V. Sharma, B. W. Gibson, S. A. Carr, O. Vitek, M. J. MacCoss, and B. MacLean. A framework for installable external tools in Skyline. *Bioinformatics*, 30(17):2521–2523, 2014.

[168] H. L. Rost, G. Rosenberger, P. Navarro, L. Gillet, S. M. Miladinovic, O. T. Schubert, W. Wolski, B. C. Collins, J. Malmstrom, L. Malmstrom, and R. Aebersold. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nature Biotechnology*, 32(3):219–223, 2014.

[169] S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*, 17(10):994–999, 1999.

[170] P. L. Ross, Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, and D. J. Pappin. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics*, 3 (12):1154–1169, 2004.

[171] A. Thompson, J. Schafer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, R. Johnstone, A. K. Mohammed, and C. Hamon. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical Chemistry*, 75 (8):1895–1904, 2003.

[172] P. J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, and A. J. Heck. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature Protocols*, 4(4):484–494, 2009.

[173] S. Y. Ow, M. Salim, J. Noirel, C. Evans, I. Rehman, and P. C. Wright. iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *Journal of Proteome Research*, 8(11):5347–5355, 2009.

[174] T. E. Thingholm, G. Palmisano, F. Kjeldsen, and M. R. Larsen. Undesirable charge-enhancement of isobaric tagged phosphopeptides leads to reduced identification efficiency. *Journal of Proteome Research*, 9(8):4045–4052, 2010.

[175] A. Sandberg, R. M. Branca, J. Lehtio, and J. Forshed. Quantitative accuracy in mass spectrometry based proteomics of complex samples: the impact of labeling and precursor interference. *Journal of Proteome Research*, 16(96):133–144, 2014.

[176] L. Ting, R. Rad, S. P. Gygi, and W. Haas. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature Methods*, 8(11):937–940, 2011.

[177] S-E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey, and M. Mann. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics*, 1(5):376–386, 2002.

[178] A. E. Merrill, A. S. Hebert, M. E. MacGilvray, C. M. Rose, D. J. Bailey, J. C. Bradley, W. W. Wood, M. El Masri, M. S. Westphall, and J. J. Gasch, A. P. Coon. NeuCode labels for relative protein quantification. *Molecular & Cellular Proteomics*, 13(9):2503–2512, 2014.

[179] T. Geiger, J. Cox, P. Ostasiewicz, J. R. Wisniewski, and M. Mann. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature Methods*, 7(5):383–385, 2010.

[180] M. Kruger, M. Moser, S. Ussar, I. Thievessen, C. A. Luber, F. Forner, S. Schmidt, S. Zanivan, R. Fassler, and M. Mann. SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function. *Cell*, 134(2):353–364, 2008.

[181] W. M. Old, K. Meyer-Arendt, L. Aveline-Wolf, K. G. Pierce, A. Mendoza, J. R. Sevinksy, K. A. Resing, and N. G. Ahn. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Molecular & Cellular Proteomics*, 4(10):1487–1502, 2005.

[182] D. H. Lundgren, S. I. Hwang, L. Wu, and D. K. Han. Role of spectral counting in quantitative proteomics. *Expert Reviews in Proteomics*, 7(1):39–53, 2010.

[183] P. Lu, C. Vogel, R. Wang, X. Yao, and E. M. Marcotte. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature Biotechnology*, 7(1):39–53, 2010.

[184] R. W. Kondrat, G. A. McClusky, and R. G. Cooks. Multiple reaction monitoring in mass spectrometry/mass spectrometry for direct analysis of complex mixtures. *Analytical Chemistry*, 50(14):2017–2021, 1978.

[185] V. Lange, P. Picotti, B. Domon, and R. Aebersold. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular Systems Biology*, 4(222), 2008.

[186] A. C. Peterson, J. D. Russell, D. J. Bailey, M. S. Westphall, and J. J. Coon. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular & Cellular Proteomics*, 11(11):1475–1488, 2012.

[187] F. Yang, N. Jaitly, H. Jayachandra, Q. Luo, M. E. Monroe, X. Du, M. A. Gritsenko, R. Zhang, D. J. Anderson, S. O. Purvine, J. N. Adkins, R. J. Moore, H. M. Mottaz, S. J. Ding, M. S. Lipton, D. G. II. Camp, H. R. Udseth, R. D. Smith, and S. Rossie. Applying a targeted label-free approach using LC-MS AMT tags to evaluate changes in protein phosphorylation following phosphatase inhibition. *Journal of Proteome Research*, 6(11):4489–4497, 2007.

[188] P. R. Cutillas and B. Vanhaesebroeck. Quantitative profile of five murine core proteomes using label-free functional proteomics. *Molecular & Cellular Proteomics*, 6(9):1560–1573, 2007.

[189] L. Beltran, P. Casado, J-C. Rodríguez-Prados, and P. R. Cutillas. Global profiling of protein kinase activities in cancer cells by mass spectrometry. *Journal of Proteomics*, 77:492–503, 2012.

[190] P. Casado and P. R. Cutillas. A self-validating quantitative mass spectrometry method for assessing the accuracy of high-content phosphoproteomic experiments. *Molecular & Cellular Proteomics*, 10(1), 2011.

[191] M. P. Alcolea, P. Casado, J. C. Rodríguez-Prados, B. Vanhaesebroeck, and P. R. Cutillas. Phosphoproteomic analysis of leukemia cells under basal and drug-treated conditions identifies markers of kinase pathway activation and mechanisms of resistance. *Molecular & Cellular Proteomics*, 11(8):453–466, 2012.

[192] V. Rajeeve, W. Pearce, M. Cascante, B. Vanhaesebroeck, and P. R. Cutillas. Polyamine production is downstream and upstream of oncogenic PI3K signalling and contributes to tumour cell growth. *Biochemical Journal*, 450(3):619–628, 2013.

[193] V. Rajeeve, I. Vendrell, E. Wilkes, N. Torbett, and P. R. Cutillas. Cross-species proteomics reveals specific modulation of signaling in cancer and stromal cells by phosphoinositide 3-kinase (PI3K) inhibitors. *Molecular & Cellular Proteomics*, 13(6):1457–1470, 2014.

[194] C. L. Han, J. S. Chen, E. C. Chan, C. P. Wu, K. H. Yu, K. T. Chen, C. C. Tsou, C. F. Tsai, C. W. Chien, Y. B. Kuo, P. Y. Lin, J. S. Yu, C. Hsueh, M. C. Chen, C. C. Chan, Y. S. Chang, and Y. J. Chen. An informatics-assisted label-free approach for personalized tissue membrane proteomics: case study on colorectal cancer. *Molecular & Cellular Proteomics*, 10(4):M110.003087, 2011.

[195] J. Cox, M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction termed MaxLFQ. *Molecular & Cellular Proteomics*, 13(9):2513–2126, 2014.

[196] B. MacLean, D. M. Tomazela, N. Schulman, M. Chambers, G. L. Finney, B. Frewen, R. Kern, D. L. Tabb, D. C. Liebler, and M. J. MacCoss. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Annual Review of Pharmacology and Toxicology*, 26(7):966–968, 2010.

[197] K. Kito and T. Ito. Mass spectrometry-based approaches toward absolute quantitative proteomics. *Current Genomics*, 9(4):263–274, 2008.

[198] I. A. Brewis and P. Brennan. Proteomics technologies for the global identification and quantification of proteins. *Advances in Protein Chemistry and Structural Biology*, 80:1–44, 2010.

[199] L. Arike, K. Valgepea, L. Peil, R. Nahku, K. Adamberg, and R. Vilu. Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli. *Journal of Proteomics*, 75(17):5437–5448, 2012.

[200] J. R. Wisniewski, M. Y. Hein, J. Cox, and M. Mann. A 'proteomic ruler' for protein copy number and concentration estimation without spike-in standards. *Molecular & Cellular Proteomics*, page mcp.M113.037309, 2014.

[201] S. A. Gerber, J. Rush, O. Stemman, M. W. Kirschner, and S. P. Gygi. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences of the United States of America*, 100(12):6940–6945, 2003.

[202] V. Brun, A. Dupuis, A. Adrait, M. Marcellin, D. Thomas, M. Court, F. Vandenesch, and J. Garin. Isotope-labeled protein standards: towards absolute quantitative proteomics. *Molecular & Cellular Proteomics*, 6(12):2139–2149, 2007.

[203] J. Malmstrom, M. Beck, A. Schmidt, V. Lange, E. W. Deutsch, and R. Aebersold. Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature*, 460(7256):762–765, 2009.

[204] R. J. Beynon, M. K. Doherty, J. M. Pratt, and S. J. Gaskell. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nature Methods*, 2(8):587–589, 2005.

[205] J. M. Pratt, D. M. Simpson, M. K. Doherty, J. Rivers, S. J. Gaskell, and R. J. Beynon. Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nature Methods*, 1(2):1029–1043, 2006.

[206] M. Uhlen, E. Bjorling, C. Agaton, C. A. Szigyarto, B. Amini, E. Andersen, A. C. Andersson, P. Angelidou, A. Asplund, C. Asplund, L. Berglund, K. Bergstrom, H. Brumer, D. Cerjan, M. Ekstrom, A. Elobeid, C. Eriksson, L. Fagerberg, R. Falk, J. Fall, M. Forsberg, M. G. Bjorklund, K. Gumbel, A. Halimi, I. Hallin, C. Hamsten, M. Hansson, M. Hedhammar, G. Hercules, C. Kampf, K. Larsson, M. Lindskog, W. Lodewyckx, J. Lund, J. Lundeberg, K. Magnusson, E. Malm, P. Nilsson, J. Odling, P. Oksvold, I. Olsson, E. Oster, J. Ottosson, L. Paavilainen, A. Persson, R. Rimini, J. Rockberg, M. Runeson, A. Sivertsson, A. Skollermo, J. Steen, M. Stenvall, F. Sterky, S. Stromberg, M. Sundberg, H. Tegel, S. Tourle, E. Wahlund, A. Walden, J. Wan, H. Wernerus, J. Westberg, K. Wester, U. Wrethagen, L. L. Xu, S. Hober, and F. Ponten. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Molecular & Cellular Proteomics*, 4(12):1920–1932, 2005.

[207] M. Zeiler, W. L. Straube, E. Lundberg, M. Uhlen, and M. Mann. A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Molecular & Cellular Proteomics*, 11(3): O111.009613, 2011.

[208] T. Maier, M. Guell, and L. Serrano. Correlation of mRNA and protein in complex biological samples. *Febs Letters*, 583(24):3966–3973, 2009.

[209] M. W. Schmidt, A. Houseman, A. R. Ivanov, and D. A. Wolf. Comparative proteomic and transcriptomic profiling of the fission yeast Schizosaccharomyces pombe. *Molecular Systems Biology*, 3:79, 2007.

[210] D. Greenbaum, C. Colangelo, K. Williams, and M. Gerstein. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology*, 4(9), 2003.

[211] S. Ghaemmaghami, W. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea, and J. S. Weissman. Global analysis of protein expression in yeast. *Nature*, 425(6959): 737–741, 2003.

[212] A. E Speers and B. F. Cravatt. Profiling enzyme activities in vivo using click chemistry methods. *Chemistry and Biology*, 11(4):535–546, 2004.

[213] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. The protein kinase complement of the human genome. *Science*, 298(48):1912–1934, 2002.

[214] B. Huang and H-Y. Kim. Interdomain conformational changes in Akt activation revealed by chemical cross-linking and tandem mass spectrometry. *Molecular & Cellular Proteomics*, 5 (6):1045–1053, 2006.

[215] C. S. Tan, C. Jørgensen, and R. Linding. Roles of "junk phosphorylation" in modulating biomolecular association of phosphorylated proteins? *Cell Cycle*, 9(7):1276–1280, 2010.

[216] G. E. Leinhard. Non-functional phosphorylations? *Trends in Biochemical Sciences*, 33(8): 351–352, 2008.

[217] C. R. Landry, L. Freschi, T. Zarin, and A. M. Moses. Turnover of protein phosphorylation evolving under stabilizing selection. *Frontiers in Genetics*, 5:245, 2014.

[218] H. W. Yung, D. S. Charnock-Jones, and G. J. Burton. Regulation of AKT phosphorylation at Ser473 and Thr308 by endoplasmic reticulum stress modulates substrate specificity in a severity dependent manner. *PLoS ONE*, 6(3):e17894, 2011.

[219] A. Schenck, L. Goto-Silva, C. Collinet, M. Rhinn, A. Giner, B. Habermann, M. Brand, and M. Zerial. The endosomal protein Appl1 mediates Akt substrate specificity and cell survival in vertebrate development. *Cell*, 133(3):486–497, 2008.

[220] M. F. Byford. Rapid and selective modification of phosphoserine residues catalysed by Ba2+ ions for their detection during peptide microsequencing. *Biochemical Journal*, 280(1): 261–265, 1991.

[221] D. T. McLachlin and B. T. Chait. Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Analytical Chemistry*, 75(24):6826–6836, 2003.

[222] M. Adamcyzk, J. C. Gelber, and J. Wu. Selective analysis of phosphopeptides within a protein mixture by chemical modification, reversible biotinylation and mass spectrometry. *Rapid Communications in Mass Spectrometry*, 15(16):1481–1488, 2001.

[223] Y. Oda, T. Nagasu, and B. Chait. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nature Biotechnology*, 19:379–382, 2001.

[224] G. Han, M. Ye, H. Zhou, X. Jiang, S. Feng, X. Jiang, R. Tian, D. Wan, H. Zou, and J. Gu. Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. *Proteomics*, 8(7): 1346–1361, 2008.

[225] D. E. McNulty and R. S. Annan. Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Molecular & Cellular Proteomics*, 7(5):971–980, 2008.

[226] T. S. Batth, C. Francavilla, and J. V. Olsen. Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics. *Journal of Proteome Research*, 2014.

[227] M. Dong, M. Ye, K. Cheng, C. Song, Y. Pan, C. Wang, Y. Bian, and H. Zou. Depletion of acidic phosphopeptides by SAX to improve the coverage for the detection of basophilic kinase substrates. *Journal of Proteome Research*, 11(9):4673–4681, 2012.

[228] M. Zarei, A. Sprenger, F. Metzger, C. Gretzmeier, and J. Dengjel. Comparison of ERLIC-TiO2, HILIC-TiO2, and SCX-TiO2 for global phosphoproteomics approaches. *Journal of Proteome Research*, 10(8):3474–3483, 2011.

[229] H. Zhou, S. Di Palma, C. Preisinger, M. Peng, A. N. Polat, A. J. Heck, and S. Mohammed. Toward a comprehensive characterization of a human cancer cell phosphoproteome. *Journal of Proteome Research*, 12(1):260–271, 2013.

[230] N. A. Patel, A. Crombie, S. E. Slade, K. Thalassinos, C. Hughes, J. B. Connolly, J. Langridge, J. C. Murrell, and J. H. Scrivens. Comparison of one- and two-dimensional liquid chromatography approaches in the label-free quantitative analysis of *Methylocella silvestris*. *Journal of Proteome Research*, 11(9):4755–4763, 2012.

[231] S. Magdeldin, J. J. Moresco, T. Yamamoto, and J. R. Yates III. Off-Line Multidimensional Liquid Chromatography and Auto Sampling Result in Sample Loss in LC/LC–MS/MS. *Journal of Proteome Research*, 13(8):3826–3836, 2014.

[232] M. Gronborg, T. Z. Kristiansen, A. Stensballe, J. S. Andersen, O. Ohara, M. Mann, O. N. Jensen, and A. Pandey. A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Molecular & Cellular Proteomics*, 1(7):517–527, 2002.

[233] J. V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, and M. Mann. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*, 127(3):635–648, 2006.

[234] B. Blagoev, S-E. Ong, I. Kratchmarova, and M. Mann. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nature Biotechnology*, 22(9):1139–1145, 2004.

[235] G. Zhang and T. A. Neubert. Use of detergents to increase selectivity of immunoprecipitation of tyrosine phosphorylated peptides prior to identification by MALDI quadrupole-TOF MS. *Proteomics*, 6(2):571–578, 2006.

[236] J. Rush, A. Mortiz, K. A. Lee, A. Guo, V. L. Goss, E. J. Spek, H. Zhang, X. M. Zha, R. D. Polakiewicz, and M. J. Comb. Use of detergents to increase selectivity of immunoprecipitation of tyrosine phosphorylated peptides prior to identification by MALDI quadrupole-TOF MS. *Nature Biotechnology*, 23(1):94–101, 2005.

[237] M. W. Pinkse, P. M. Uitto, M. J. Hilhorst, B. Ooms, and A. J. Heck. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Analytical Chemistry*, 76(14):3935–3943, 2004.

[238] U. K. Aryal and A. R. S. Ross. Enrichment and analysis of phosphopeptides under different experimental conditions using titanium dioxide affinity chromatography and mass spectrometry. *Rapid Communications in Mass Spectrometry*, 24(2):219–231, 2009.

[239] M. R. Larsen, T. E. Thingholm, O. N. Jensen, P. Roepstorff, and T. J. Jorgensen. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Molecular & Cellular Proteomics*, 4(7):873–886, 2005.

[240] A. N. Kettenbach and S. A. Gerber. Rapid and reproducible single-stage phosphopeptide enrichment of complex peptide mixtures: Application to general and phosphotyrosine-specific phosphoproteomics experiments. *Analytical Chemistry*, 83(70):7635–7644, 2011.

[241] K. Engholm-Keller, P. Birck, J. Storling, F. Pociot, T. Mandrup-Poulsen, and M. R. Larsen. TiSH – a robust and sensitive global phosphoproteomics strategy employing a combination of TiO2, SIMAC, and HILIC. *Journal of Proteomics*, 75(18):5749–5761, 2012.

[242] G. Muszynska, G. Gobrowolska, A. Medin, P. Ekman, and J. O. Porath. Model studies on iron(III) ion affinity chromatography. II. Interaction of immobilized iron(III) ions with phosphorylated amino acids, peptides and proteins. *Journal of Chromatography*, 604(1):19–28, 1992.

[243] S. Li and C. Dass. Iron(III)-immobilized metal ion affinity chromatography and mass spectrometry for the purification and characterization of synthetic phosphopeptides. *Analytical Biochemistry*, 270(1):9–14, 1999.

[244] E. L. de Graff, P. Giansanti, A. F. Maarten Altelaar, and Heck. A. J. Single-step enrichment by Ti4+-IMAC and label-free quantitation enables in-depth monitoring of phosphorylation dynamics with high reproducibility and temporal resolution. *Molecular & Cellular Proteomics*, 13(9):2426–2434, 2014.

[245] C. F. Tsai, Y. T. Wang, Y. R. Chen, C. Y. Lai, P. Y. Lin, K. T. Pan, J. Y. Chen, K. H. Khoo, and Y. J. Chen. Immobilized metal affinity chromatography revisited: pH/acid control toward high selectivity in phosphoproteomics. *Journal of Proteome Research*, 7(9):4058–4069, 2008.

[246] J. Ye, X. Zhang, C. Young, X. Zhao, Q. Hao, L. Cheng, and O. N. Jensen. Optimized IMAC-IMAC protocol for phosphopeptide recovery from complex biological samples. *Journal of Proteome Research*, 9(7):3561–3573, 2010.

[247] T. E. Thingholm, O. N. Jensen, P. J. Robinson, and M. R. Larsen. SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Molecular & Cellular Proteomics*, 7(4):661–671, 2008.

[248] M. P. Jedrychowski, E. L. Huttlin, W. Haas, M. E. Sowa, R. Rad, and S. P. Gygi. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Molecular & Cellular Proteomics*, 10(12):M111.009910, 2011.

[249] C. K. Freze, A. F. M. Altelaar, H. van den Toorn, D. Nolting, J. Griep-Raming, and A. J. Heck. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and high-energy collision dissociation tandem mass spectrometry. *Analytical Chemistry*, 84 (22):9668–9673, 2012.

[250] C. K. Freze, H. Zhou, T. Taus, A. F. M. Altelaar, K. Mechtler, A. J. Heck, and S. Mohammed. Unambiguous phosphosite localisation using electron-transfer/higher-energy collision dissociation (EThcD). *Journal of Proteome Research*, 12(3):1520–1525, 2013.

[251] D. L. Swaney, G. C. McAlister, and J. J. Coon. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nature Methods*, 5:959–964, 2008.

[252] M. M. Savitski, S. Lemeer, M. Boesche, M. Lang, T. Mathieson, M. Bantscheff, and B. Kuster. Confident phosphorylation site localisation using the mascot delta score. *Molecular & Cellular Proteomics*, 10(2):M110.003830, 2011.

[253] T. Taus, T. Kocher, P. Pichler, C. Paschke, A. Schmidt, C. Henrich, and K. Mechtler. Universal and confident phosphorylation site localization using PhosphoRS. *Journal of Proteome Research*, 10(12):5354–5362, 2011.

[254] D. Fermin, S. J. Walmsley, A. C. Gingras, H. Choi, and A. I. Nesvizhskii. LuciPHOr: algorithm for phosphorylation site localization with false localization rate estimation using modified target-decoy approach. *Molecular & Cellular Proteomics*, 12(11):3409–3419, 2008.

[255] *R: A language and environment for statistical computing*, 2014.

[256] H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York, 2009.

[257] H. Wickham. Reshaping data with the reshape package. *Journal of Statistical Software*, 21:1–20, 2007.

[258] K. Soetaert. *plot3D: Plotting multi-dimensional data*, 2014. R package version 1.0-1.

[259] D. T. Lang. *XMLRPC: Remote Procedure Call (RPC) via XML in R*. R package version 0.3-0.

[260] P. Shannon. *RCytoscape: Display and manipulate graphs in Cytoscape*, 2013.

[261] G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.

[262] G. K. Smyth. *Limma: linear models for microarray data*, pages 397–420. Springer New York, 2005.

[263] G. R. Warnes, B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw, T. Lumley, M. Maechler, A. Magnusson, S. Moeller, M. Schwartz, and B. Venables. *gplots: Various R programming tools for plotting data*, 2013.

[264] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, 2003.

[265] B. M. Bolstad, R. A. Irizarry, M. Astrand, and T. P. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2):185–193, 2003.

[266] G. K. Smyth. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article3, 2004.

[267] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57(1):289–300, 1995.

[268] T. E. Thingholm, O. N. Jensen, and M. R. Larsen. Analytical strategies for phosphoproteomics. *Proteomics*, 9(6):1451–1468, 2009.

[269] M. S. Sachdeva and J. R. Babu. *Chromatography*, page 620. Lippincott Williams & Wilkins, 2006.

[270] J. R. Griffiths, S. Perkins, Y. Connolly, L. Zhang, M. Holland, V. Barattini, L. Pereira, A. Edge, H. Ritchie, and D. L. Smith. The utility of porous graphitic carbon as a stationary phase in proteomics workflows: Two-dimensional chromatography of complex peptide samples. *Journal of Chromatography A*, 1232:276–280, 2003.

[271] N. Morrice. EuPA/BSPR conference, Glasgow, 2012. Personal Communication.

[272] A. Zerck, E. Nordhoff, A. Resemann, E. Mirgorodskaya, D. Suckau, K. Reinert, H. Lehrach, and J. Gobom. An iterative strategy for precursor ion selection for LC-MS/MS based shotgun proteomics. *Journal of Proteome Research*, 8(7):3239–3251, 2009.

[273] A. Scherl, S. A. Shaffer, G. K. Taylor, H. D. Kulasekara, S. I. Miller, and D. R. Goodlett. Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides. *Journal of Proteome Research*, 80(4):1182–1191, 2008.

[274] D. L. Tabb, L. Vega-Montoto, P. A. Rudnick, A. M. Variyath, L. A-J. Ham, D. M. Bunk, L. E. Kilpatrick, D. Billheimer, R. K. Blackman, H. L. Cardasis, S. A. Carr, K. R. Clauser, J. D. Jaffe,

K. A. Kowalski, T. A. Neubert, F. E. Regnier, B. Schilling, T. J. Tegeler, M. Wang, P. Wang, J. R. Whiteaker, L. J. Zimmerman, S. J. Fisher, B. W. Gibson, C. R. Kinsinger, M. Mesri, H. Rodriguez, S. E. Stein, P. Tempst, A. G. Paulovich, D. C. Liebler, and C. Speigelman. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *Journal of Proteome Research*, 9(2):761–776, 2010.

[275] P. J. Ulintz, A. K. Yocum, B. Bodenmiller, R. Aebersold, P. C. Andrews, and A. I. Nesvizhskii. Comparison of MS2-only, MSA, and MS2/MS3 methodologies for phosphopeptide identification. *Journal of Proteome Research*, 8(2):887–899, 2008.

[276] A. B. Caldwell, Z. Cheng, J. D. Vargas, H. A. Birnbaum, and A. Hoffmann. Network dynamics determine the autocrine and paracrine signaling functions of TNF. *Genes & Development*, 28(19):2120–2133, 2014.

[277] C. M. Tsai, H. Y. Wu, T. H. Su, C. W. Kuo, H. W. Huang, C. H. Chung, C. S. Chen, K. H. Khoo, Y. J. Chen, and K. I. Lin. Phosphoproteomic analyses reveal that galectin-1 augments the dynamics of B-cell receptor signaling. *Journal of Proteomics*, 103:241–2553, 2014.

[278] A. Perino, M. Beretta, A. Kilic, A. Ghigo, D. Carnevale, I. E. Repetto, L. Braccini, D. Longo, M. Leibig-Gonglack, T. Zaglia, R. Iacobucci, M. Mongillo, R. Wetzker, M. Bauer, S. Aime, A. Vercelli, G. Lembo, A. Preifer, and E. Hirsch. Combined inhibition of PI3Kβ and PI3Kγ reduces fat mass by enhancing α-MSH–dependent sympathetic drive. *Science Signaling*, 7(352):ra110, 2014.

[279] E. Fadhal, J. Gamieldien, and E. C. Mwanbene. Protein interaction networks as metric spaces: a novel perspective on distribution of hubs. *BMC Systems Biology*, 18(8):6, 2014.

[280] B. Bodenmiller, S. Wanka, C. Kraft, J. Urban, D. Campbell, P. G. Pedrioli, B. Gerrits, P. Picotti, H. Lam, O. Vitek, M-Y. Brusniak, C. Roschitzki, B.and Zhang, K. M. Shokat, R. Schlapbach, A. Colman-Lerner, G. P. Nolan, A. I. Nesvizhskii, M. Peter, R. Loewith, C. von Mering, and R. Aebersold. Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Science Signaling*, 3 (153), 2010.

[281] S. D. Santos, P. J. Verveer, and P. I. Bastiaens. Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nature Cell Biology*, 9(3):324–330, 2007.

[282] S. S. McMahon, A. Sim, S. Filippi, R. Johnson, J. Liepe, D. Smith, and M. P. H. Stumpf. Information theory and signal transduction systems: From molecular information processing to network inference. *Seminars in Cell & Developmental Biology*, 35C:98–108, 2014.

[283] M. Jeanmougin, A. de Reynies, L. Marisa, C. Paccard, G. Nuel, and M. Guedj. Should we abandon the *t*-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies. *PLoS ONE*, 5(9):e12336, 2010.

[284] L. Ting, M. J. Cowley, S. L. Hoon, M. Guilhaus, M. J. Raftery, and R. Cavicchioli. Normalization and statistical analysis of quantitative proteomics data generated by metabolic labeling. *Molecular & Cellular Proteomics*, 8(10):2227–2242, 2009.

[285] S. Y. Kim and D. J. Volsky. PAGE: parametric analysis of gene set enrichment. *BMC Bioinformatics*, 8(6):144, 2005.

[286] M. F. Ciaccio, J. P. Wagner, C-P. Chuu, D. A. Lauffenburger, and R. B. Jones. Systems analysis of EGF Receptor signaling dynamics with micro-western arrays. *Nature Methods*, 7(2):148–155, 2010.

[287] A. M. Vincent and E. L. Feldman. Control of cell survival by IGF signaling pathways. *Growth Hormone & IGF Research*, 12(4):193–197, 2002.

[288] L. K. Goh and A. Sorkin. Endocytosis of receptor tyrosine kinases. *Cold Spring Harbor Perspectives in Biology*, 5(5):a017459, 2013.

[289] G. Lurje and H. J. Lenz. EGFR signaling and drug discovery. *Oncology*, 77(6):400–410, 2009.

[290] F. I. Arnaldez and L. J. Helman. Targeting the insulin growth factor receptor 1. *Haematology/Oncology Clinics of North America*, 26(3):527–542, 2012.

[291] E. E. Er, M. C. Mendoz, A. M. Mackey, L. E. Rameh, and J. Blenis. AKT facilitates EGFR trafficking and degradation by phosphorylating and activating PIKfyve. *Science Signaling*, 6(279):ra45, 2013.

[292] M. R. Larsen, M. E. Graham, P. J. Robinson, and P. Roepstorff. Improved detection of hydrophilic phosphopeptides using graphite powder microcolumns and mass spectrometry: evidence for *in vivo* doubly phosphorylated dynamin I and dynamin III. *Molecular & Cellular Proteomics*, 3(5):456–465, 2004.

[293] Q-R. Li, Z-B. Ning, J-S. Tang, S. Nie, and R. Zeng. Effect of peptide-to-TiO_2 beads ratio on phosphopeptide enrichment selectivity. *Journal of Proteome Research*, 8(11):5375–5381, 2009.

[294] C. J. Tape, J. D. Worboys, J. Sinclair, R. Gourlay, J. Vogt, K. M. McMahon, M. Trost, D. A. Lauffenburger, D. J. Lamont, and C. Jørgensen. Reproducible automated phosphopeptide enrichment using magnetic TiO_2 and Ti-IMAC. *Analytical Chemistry*, 86(20):10296–10302, 2014.

[295] A. Guthals and N. Bandeira. Peptide identification by tandem mass spectrometry with alternate fragmentation modes. *Molecular & Cellular Proteomics*, 11(9):550–557, 2012.

[296] K. Oda, Y. Matsuoka, A. Funahashi, and H. Kitano. A comprehensive pathway map of epidermal growth factor receptor signaling. *Molecular Systems Biology*, 1:2005.0010, 2005.

[297] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M. Tanabe. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, 40((Database issue)):D109–114, 2012.

[298] J. E. Elias and S. P. Gygi. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods*, 4(3):207–214, 2007.

[299] X. Qui, H. Wu, and R. Hu. The impact of quantile and rank normalization procedures on the testing power of gene differential expression analysis. *BMC Bioinformatics*, 11(14):124, 2013.

[300] S. J. Callister, R. C. Barry, J. N. Adkins, E. T. Johnson, W-J. Qian, B-J. M. Webb-Robertson, R. D. Smith, and M. S. Lipton. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *Journal of Proteome Research*, 5(2):277–286, 2006.

[301] P. I. Poulikakos, C. Zhang, G. Bollag, K. M. Shokat, and N. Rosen. RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. *Nature*, 464(7287): 427430390, 2010.

[302] C. W. Lindsley, Z. Zhao, W. H. Leister, R. G. Robinson, S. F. Barnett, D. Defeo-Jones, R. E. Jones, G. D. Hartman, J. R. Huff, H. E. Huber, and M. E. Duggan. Allosteric Akt (PKB) inhibitors: discovery and SAR of isozyme selective inhibitors. *Bioorganic & and Medicinal Chemistry Letters*, 15(3):761–764, 2005.

[303] M. Sumi, K. Kiuchi, T. Ishikawa, A. Ishii, M. Hagiwara, T. Nagatsu, and H. Hidaka. The newly synthesized selective Ca2+/calmodulin dependent protein kinase II inhibitor KN-93 reduces dopamine contents in PC12h cells. *Biochemical and Biophysical Research Communications*, 181(3):968–975, 1991.

[304] H. Tokumitsu, T. Chijiwa, M. Hagiwara, A. Mizutani, M. Terasawa, and H. Hidaka. KN-62, 1-[N,O-bis(5-isoquinolinesulfonyl)-N-methyl-L-tyrosyl]-4-phenylpiperazi ne, a specific inhibitor of Ca2+/calmodulin-dependent protein kinase II. *Journal of Biological Chemistry*, 265(8): 4315–4320, 1990.

[305] C. N. Hancock, A. Macias, E. K. Lee, S. Y. Yu, A. D. Mackerell Jr, and P. Shapiro. Identification of novel extracellular signal-regulated kinase docking domain inhibitors. *Journal of Medicinal Chemistry*, 48(14):4586–4595, 2005.

[306] M. Ohori, T. Kinoshita, M. Okubo, K. Sato, A. Yamazaki, H. Arakawa, S. Nishimura, N. Inamura, H. Nakajima, M. Neya, H. Miyake, and T. Fujii. Identification of a selective ERK inhibitor and structural determination of the inhibitor-ERK2 complex. *Biochemical and Biophysical Research Communications*, 336(1):357–363, 2005.

[307] M. Tamura, H. Nakao, H. Yoshizaki, M. Shiratsuchi, H. Shigyo, H. Yamada, T. Ozawa, J. Totsuka, and H. Hidaka. Development of specific Rho-kinase inhibitors and their clinical application. *Biochimica Et Biophysica Acta*, 1754(1-2):245–252, 2005.

[308] T. Ishizaki, M. Uehata, I. Tamechika, J. Keel, K. Nonomura, M. Maekawa, and S. Narumiya. Pharmacological properties of Y-27632, a specific inhibitor of rho-associated kinases. *Molecular Pharmacology*, 57(5):976–983, 2000.

[309] K. Mahajan and N. P. Mahajan. PI3K-independent AKT activation in cancers: a treasure trove for novel therapeutics. *Journal of Cellular Physiology*, 227(9):3178–3184, 2012.

[310] K. Mahajan, D. Coppola, S. Challa, B. Fang, Y. A. Chen, Lopez A. S., J. Koomen, R. W. Engelman, C. Rivera, R. S. Muraoka-Cook, J. Q. Cheng, E. Schönbrunn, S. M. Sebti, H. S. Earp, and N. P. Mahajan. Ack1 mediated AKT/PKB tyrosine 176 phosphorylation regulates its activation. *PLoS ONE*, 5(3):e9646, 2010.

[311] Y. Zheng, M. Peng, Z. Wang, J. M. Asara, and A. L. Tyner. Protein tyrosine kinase 6 directly phosphorylates AKT and promotes AKT activation in response to epidermal growth factor. *Molecular & Cellular Biology*, 30(17):4280–4292, 2010.

[312] R. Chen, O. Kim, J. Yang, K. Sato, K. M. Esienmann, J. McCarthy, H. Chen, and Y. Qiu. Regulation of Akt/PKB activation by tyrosine phosphorylation. *Journal of Biological Chemistry*, 276(34):31858–31862, 2001.

[313] J. Moffat and D. M. Sabatini. Building mammalian signalling pathways with RNAi screens. *Nature Reviews Molecular Cellular Biology*, 7(3):177–187, 2006.

[314] J. Rauch, N. Volinsky, D. Romano, and W. Kolch. The secret life of kinases: functions beyond catalysis. *Cell Communication & Signaling*, 9(1):23, 2011.

[315] I. Sánchez-Hernández, P. Baguero, L. Calleros, and A. Chiloeches. Dual inhibition of (V600E)BRAF and the PI3K/AKT/mTOR pathway cooperates to induce apoptosis in melanoma cells through a MEK-independent mechanism. *Cancer Letters*, 312(2):244–255, 2012.

[316] A. Lassen, M. Atefi, L. Robert, D. J. Wong, M. Cerniglia, B. Comin-Anduix, and A. Ribas. Effects of AKT inhibitor therapy in response and resistance to BRAF inhibition in melanoma. *Molecular Cancer*, 13(83), 2014.

[317] A. M. Pereira, C. Tudor, P-A. Pouille, S. Shekhar, J. S. Kanger, V. Subramaniam, and E. Martín-Blanco. Plasticity of the MAPK signaling network in response to mechanical stress. *PLoS ONE*, 9(7):e101963, 2014.

[318] A. Krämer, J. Green, J. Pollard Jr, and S. Tugendreich. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*, 30(4):523–530, 2014.

[319] A. Lan, M. Ziv-Ukelson, and E. Yeger-Lotem. A context-sensitive framework for the analysis of human signalling pathways in molecular interaction networks. *Bioinformatics*, 29(13): i210–6, 2013.

[320] J. S. Duncan, M. C. Whittle, K. Nakamura, A. N. Abell, A. A. Midland, J. S. Zawistowski, N. L. Johnson, D. A. Granger, N. V. Jordan, D. B. Darr, J. Usary, P. F. Kuan, D. M. Smalley, B. Major, X. He, K. A. Hoadley, B. Zhou, N. E. Sharpless, C. M. Perou, W. Y. Kim, S. M. Gomez, X. Chen, J. Jin, S. V. Frye, H. S. Earp, L. M. Graves, and G. L. Johnson. Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer. *Cell*, 149(2):307–321, 2012.

[321] M. J. Cooper, N. J. Cox, E. I. Zimmerman, B. J. Dewar, J. S. Duncan, M. C. Whittle, T. A. Nguyen, L. S. Jones, S. Ghose Roy, D. M. Smalley, P. F. Kuan, K. L. Richards, R. I. Christopherson, J. Jin, S. V. Frye, G. L. Johnson, A. S. Baldwin, and L. M. Graves. Application of multiplexed kinase inhibitor beads to study kinome adaptations in drug-resistant leukemia. *PLoS ONE*, 8(6):e66755, 2013.

[322] J. E. Toettcher, O. D. Weiner, and W. A. Lim. Using optogenetics to interrogate the dynamic control of signal transmission by the Ras/ERK module. *Cell*, 155(6):1422–1434, 2013.

[323] R. Zielienski, P. F. Przytycki, J. Zheng, D. Zhang, T. M. Pryztycka, and J. Capala. The crosstalk between EGF, IGF, and Insulin cell signalling pathways – computational and expeirmental analysis. *BMC Systems Biology*, 3(88), 2009.

[324] T. Muranen, L. M. Selfors, D. T. Worster, M. P. Iwanicki, L. Song, F. C. Morales, S. Gao, G. B. Mills, and J. S. Brugge. Inhibition of PI3K/mTOR leads to adaptive resistance in matrix-attached cancer cells. *Cancer Cell*, 21(2):227–239, 2012.

[325] P. Liu, H. Cheng, S. Santiago, M. Raeder, F. Zhang, A. Isabella, J. Yang, D. J. Semaan, C. Chen, E. A. Fox, N. S. Gray, J. Monahan, R. Schlegel, R. Beroukhim, G. B. Mills, and

J. J. Zhao. Oncogenic PIK3CA-driven mammary tumors frequently recur via PI3K pathway-dependent and PI3K pathway-independent mechanisms. *Nature Medicine*, 17(9):1116–1120, 2011.

[326] M. Elkabets, S. Vora, D. Juric, N. Morse, M. Mino-Kenudson, T. Muranen, J. Tao, A. B. Campos, J. Rodon, Y. H. Ibrahim, V. Serra, V. Rodri-Outmezguine, S. Hazra, S. Singh, P. Kim, C. Qaudt, M. Liu, A. Huang, N. Rosen, J. A. Engelman, M. Scaltriti, and J. Baselga. mTORC1 Inhibition Is Required for Sensitivity to PI3K p110 alpha Inhibitors in PIK3CA-Mutant Breast Cancer. *Science Translational Medicine*, 5(196):196ra99, 2013.

[327] S. J. Klempner, A. P. Myers, and L. C. Cantley. What a tangled web we weave: emerging resistance mechanisms to inhibition of the phosphoinositide 3-kinase pathway. *Cancer Discovery*, 3(12):1345–1354, 2013.

[328] P. Lito, N. Rosen, and D. B. Solit. Tumor adaptation and resistance to RAF inhibitors. *Nature Medicine*, 19(11):1401–1409, 2013.

[329] M.. Geringer. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England Journal of Medicine*, 366(10):883–892, 2012.

[330] I. P. Janecka. Cancer control through principles of system science, complexity, and chaos theory: a model. *International Journal of Medical Sciences*, 4(3):164–173, 2007.

[331] F. Grizzi, A. Di Ieva, C. Russo, E. E. Frezza, E. Cobos, P. C. Muzzio, and M. Chiriva-Internati. Cancer initiation and progression: an unsimplifiable complexity. *Theoretical Biology and Medical Modelling*, 3:37, 2006.

[332] B. P. Bezruchko and D. A. Smirnov. *Extracting knowledge from time series: an introduction to nonlinear empirical modeling*. Springer, 2010.

[333] P. Bryant, R. Brown, and H. D. Abarbanel. Lyapunov exponents from observed time series. *Physical Review Letters*, 65(13):1523–1526, 1990.

Appendix 1

Summary of CTAM identifications (Chapter 4).

CTAM group	Phosphorylation site
Akt-CAMK2-EGFR-ERK-MEK-mTOR	PHF12(S977)
Akt-CAMK2-EGFR-ERK-MEK-mTOR	PRCC(S267)
Akt-CAMK2-EGFR-ERK-MEK-mTOR	SERPINA10(T297)
Akt-CAMK2-EGFR-ERK-MEK-mTOR	SSFA2(S737+S739)
Akt-CAMK2-EGFR-ERK-MEK	EIF3G(T41+S42)
Akt-CAMK2-EGFR-ERK-MEK	JUND(S90)
Akt-CAMK2-EGFR-ERK-MEK	PBK(T24+S32)
Akt-CAMK2-EGFR-ERK-MEK	RPLP2(S102+S105)
Akt-CAMK2-EGFR-ERK-ROCK	ANXA13(Y291+S294)
Akt-CAMK2-ERK-MEK-mTOR	USP13(S630)
Akt-CAMK2-mTOR-p70S6K-PI3K	RICTOR(T1135+S1138)
Akt-ERK-mTOR-p70S6K-PI3K	ZNF687(S1118)
Akt-mTOR-p70S6K-PI3K-PKC	WWTR1(S93)
Akt-CAMK2-EGFR-ERK	DSP(S38)
Akt-CAMK2-EGFR-ERK	KAT7(T302+S303)
Akt-CAMK2-EGFR-ERK	KLF16(S230)
Akt-CAMK2-EGFR-ERK	NOS1AP(S371+S374)
Akt-CAMK2-EGFR-ERK	PER2(S77)
Akt-CAMK2-EGFR-ERK	TNKS1BP1(S494+S504)
Akt-CAMK2-EGFR-mTOR	PLEC(T4030)
Akt-CAMK2-EGFR-p70S6K	GAPDH(T184)
Akt-CAMK2-ERK-MEK	FBRSL1(T989+T1010)
Akt-CAMK2-mTOR-PI3K	EIF4EBP1(T41+S44)
Akt-EGFR-ERK-mTOR	DLGAP4(S973)
Akt-mTOR-p70S6K-PI3K	C10orf47(S43)
Akt-mTOR-p70S6K-PI3K	C2orf78(S624+S625)
Akt-mTOR-p70S6K-PI3K	CHD1(S1385)
Akt-mTOR-p70S6K-PI3K	CHD1(S1387)
Akt-mTOR-p70S6K-PI3K	CIC(S299)
Akt-mTOR-p70S6K-PI3K	CLMN(S921)
Akt-mTOR-p70S6K-PI3K	CNBP(S47)
Akt-mTOR-p70S6K-PI3K	CTDSPL2(S28)
Akt-mTOR-p70S6K-PI3K	EHMT2(S232)
Akt-mTOR-p70S6K-PI3K	EPRS(T956)
Akt-mTOR-p70S6K-PI3K	GTPBP4(T201+T202)
Akt-mTOR-p70S6K-PI3K	HMGCR(S872)
Akt-mTOR-p70S6K-PI3K	IFT122(Y146+S157)
Akt-mTOR-p70S6K-PI3K	KRT8(S477)
Akt-mTOR-p70S6K-PI3K	MACF1(S1376)
Akt-mTOR-p70S6K-PI3K	MACF1(S1378)
Akt-mTOR-p70S6K-PI3K	MTDH(S494)
Akt-mTOR-p70S6K-PI3K	MYO5A(S1652)
Akt-mTOR-p70S6K-PI3K	NEK9(S332)
Akt-mTOR-p70S6K-PI3K	NOLC1(S686+S698)
Akt-mTOR-p70S6K-PI3K	PCBP2(S272)
Akt-mTOR-p70S6K-PI3K	PLEC(S4406)
Akt-mTOR-p70S6K-PI3K	PLEC(T4411)
Akt-mTOR-p70S6K-PI3K	RBL2(S1138)
Akt-mTOR-p70S6K-PI3K	RBM14(S618)
Akt-mTOR-p70S6K-PI3K	RICTOR(T1135)
Akt-mTOR-p70S6K-PI3K	RPS6KB1(S427)
Akt-mTOR-p70S6K-PI3K	SART1(S448)
Akt-mTOR-p70S6K-PI3K	SETD1A(S1171)

Akt-mTOR-p70S6K-PI3K	SPATS2L(S135)
Akt-mTOR-p70S6K-PI3K	UCK1(S253)
Akt-mTOR-p70S6K-PI3K	WDHD1(S374)
Akt-mTOR-p70S6K-PI3K	YBX1(S209)
Akt-mTOR-p70S6K-PI3K	ZFP36L1(S334)
p70S6K-PI3K-PKC-ROCK	ZYX(S143)
Akt-CAMK2-EGFR	ATG4B(S383)
Akt-CAMK2-EGFR	DSP(T2612)
Akt-CAMK2-EGFR	FAM177A1(S70+T71)
Akt-CAMK2-EGFR	LSR(S643+S646)
Akt-CAMK2-EGFR	MAP7D1(S834)
Akt-CAMK2-EGFR	NCOR2(S1018)
Akt-CAMK2-EGFR	NOTCH2(S1804)
Akt-CAMK2-EGFR	PTGES3(S148)
Akt-CAMK2-EGFR	RANBP2(S2628)
Akt-CAMK2-EGFR	RANGAP1(S428+S442)
Akt-CAMK2-EGFR	ROCK2(S1134+S1137)
Akt-CAMK2-EGFR	SMARCAD1(S124+S127)
Akt-CAMK2-EGFR	SRRM2(S1398+S1404)
Akt-CAMK2-EGFR	TRMT112(S119)
Akt-CAMK2-EGFR	YWHAQ(S230)
Akt-CAMK2-ERK	GORASP2(T423+S451)
Akt-CAMK2-ERK	PRKCD(T507)
Akt-EGFR-ERK	AHNAK(S1298)
Akt-EGFR-ERK	MAP2K4(T391+S394)
Akt-EGFR-ERK	MAVS(S258)
Akt-EGFR-ERK	MCM3(S711+T722)
Akt-EGFR-MEK	RSF1(S1359+S1375)
Akt-ERK-MEK	PPP1R12A(T696)
Akt-mTOR-PI3K	AKT1S1(S183)
Akt-mTOR-PI3K	ARFGAP2(S368)
Akt-mTOR-PI3K	BAHD1(S121)
Akt-mTOR-PI3K	BCL7C(S126)
Akt-mTOR-PI3K	CCNL1(S335+S338)
Akt-mTOR-PI3K	CTTN(T440)
Akt-mTOR-PI3K	EEF2K(S71+S74)
Akt-mTOR-PI3K	EIF4EBP1(S65+T68)
Akt-mTOR-PI3K	EIF4EBP1(T37+T41)
Akt-mTOR-PI3K	EIF4EBP1(T68+T70)
Akt-mTOR-PI3K	EIF4EBP2(S65+T70)
Akt-mTOR-PI3K	EIF4EBP2(T70)
Akt-mTOR-PI3K	EIF4EBP2(Y34+T46)
Akt-mTOR-PI3K	FAM54A(S305)
Akt-mTOR-PI3K	FAM83H(S1024)
Akt-mTOR-PI3K	FOXK1(S249+S253)
Akt-mTOR-PI3K	FOXK1(S257)
Akt-mTOR-PI3K	FOXK2(S369)
Akt-mTOR-PI3K	FOXK2(S385)
Akt-mTOR-PI3K	GSK3B(S9)
Akt-mTOR-PI3K	LAMTOR1(S98)
Akt-mTOR-PI3K	LARP1(S1058)
Akt-mTOR-PI3K	LARP1(S774)
Akt-mTOR-PI3K	LIMA1(S230)
Akt-mTOR-PI3K	MYC(S62)
Akt-mTOR-PI3K	MYCBP2(S2833)
Akt-mTOR-PI3K	PCBP1(S246)
Akt-mTOR-PI3K	PCBP1(S262)
Akt-mTOR-PI3K	PCBP1(S263)

Akt-mTOR-PI3K	PDE3B(S494)
Akt-mTOR-PI3K	PUM1(S209)
Akt-mTOR-PI3K	RITA(S248)
Akt-mTOR-PI3K	RN3L1(S44)
Akt-mTOR-PI3K	RPS6KB1(S447+S452)
Akt-mTOR-PI3K	SEC14L1(T234)
Akt-mTOR-PI3K	SUN2(S20)
Akt-mTOR-PI3K	TBX2(S657)
Akt-mTOR-PI3K	TP53BP1(S1430)
Akt-mTOR-PI3K	TPR(T650)
Akt-mTOR-PI3K	UBA1(S816)
Akt-mTOR-PI3K	USP32(S1361)
Akt-mTOR-PI3K	USP32(S1366+S1371)
Akt-mTOR-PI3K	VWA7(S305+S325)
Akt-mTOR-PI3K	WDHD1(S374+S383)
Akt-mTOR-PI3K	ZNF185(S64)
EGFR-ERK-MEK	INCENP(T145+S148)
EGFR-PI3K-PKC	SPTBN1(S2358)
mTOR-p70S6K-PI3K	BOD1L1(S635)
mTOR-p70S6K-PI3K	BOD1L1(S637)
mTOR-p70S6K-PI3K	C14orf43(T698+S700+S718)
mTOR-p70S6K-PI3K	C5orf30(S167)
mTOR-p70S6K-PI3K	F12(S335+T337)
mTOR-p70S6K-PI3K	RBM14(S620)
mTOR-p70S6K-PI3K	WWTR1(S89)
p70S6K-PI3K-PKC	ZYX(S142)
Akt-CAMK2	ARHGAP35(S975+S985)
Akt-CAMK2	CACNB1(T217+S228+Y242)
Akt-CAMK2	CARHSP1(S30+S32+S41)
Akt-CAMK2	DDHD1(T726+S727)
Akt-CAMK2	EEF1B2(S106)
Akt-CAMK2	FAM122B(S222+S245)
Akt-CAMK2	FBXW7(S396)
Akt-CAMK2	FKBP15(S1114)
Akt-CAMK2	FOXO3(S75)
Akt-CAMK2	FSBP(S191+S193+T204)
Akt-CAMK2	GIT2(S418+S421)
Akt-CAMK2	GPKOW(S471)
Akt-CAMK2	INCENP(T239)
Akt-CAMK2	KIAA0284(S951+S954)
Akt-CAMK2	KIAA1429(S138)
Akt-CAMK2	KLHDC4(S413+S418)
Akt-CAMK2	MAP2K4(S257)
Akt-CAMK2	MARK4(T214)
Akt-CAMK2	MAVS(S152+T163)
Akt-CAMK2	MKL2(S207+T227)
Akt-CAMK2	NKX6-3(T8)
Akt-CAMK2	PARP4(T1119)
Akt-CAMK2	PRKACA(T198)
Akt-CAMK2	PRKACA(T202)
Akt-CAMK2	PRKCD(T511)
Akt-CAMK2	PRKCI(T410)
Akt-CAMK2	RAB11FIP2(S150)
Akt-CAMK2	RPUSD1(S271)
Akt-CAMK2	RRM2(S20)
Akt-CAMK2	SAP30BP(S18+S22)
Akt-CAMK2	SERHL2(T272)
Akt-CAMK2	SMN1(S28+S31)

Akt-CAMK2	STRN(S245)
Akt-CAMK2	TFG(S193+S197)
Akt-CAMK2	TOP2B(S1522+S1524)
Akt-CAMK2	TRIP12(S1317+S1322)
Akt-CAMK2	UBAP2L(S428)
Akt-CAMK2	UNC13D(S150)
Akt-CAMK2	WAPAL(S77)
Akt-EGFR	DCAF5(S648)
Akt-EGFR	EIF4G1(T205+T207)
Akt-EGFR	GAPVD1(S1105)
Akt-EGFR	KDM1A(T59)
Akt-EGFR	MFAP1(T267)
Akt-EGFR	PA2G4(S2+T11)
Akt-EGFR	PFDN4(S125)
Akt-EGFR	PXN(S126+S130)
Akt-EGFR	TNFRSF10A(S466)
Akt-ERK	AGAP2(S808)
Akt-mTOR	EP400(T1738)
Akt-mTOR	NSD1(S979)
Akt-mTOR	RAPGEF6(S1070)
Akt-mTOR	SERBP1(S392)
Akt-mTOR	UBA1(S816+S820)
Akt-p70S6K	BTF3(S161)
Akt-p70S6K	RAD18(S99)
Akt-PI3K	CCNL1(T325+S341+S342)
Akt-PI3K	CDC42EP2(S109)
Akt-PI3K	FOXK1(S236+S257)
Akt-PI3K	GAPVD1(S566+S569)
Akt-PI3K	SEMA4B(S825)
Akt-PI3K	STIP1(S16)
Akt-PKC	YEATS2(T1149+T1165)
CAMK2-EGFR	NUP153(S209)
CAMK2-EGFR	SCRIB(S1306+S1309)
CAMK2-EGFR	SPICE1(S760+S764)
CAMK2-EGFR	TRIM28(S43)
EGFR-ERK	C14orf43(S700+S709)
EGFR-MEK	MAPK1(T185+Y187)
EGFR-MEK	RANBP2(S1869)
EGFR-MEK	TPR(S2155)
EGFR-p70S6K	TMCC1(S382)
ERK-mTOR	C14orf43(S461)
ERK-p70S6K	RXRA(T23)
MEK-PKC	ERF(S532)
mTOR-p70S6K	USP5(S785)
mTOR-PI3K	ATRX(S594)
mTOR-PI3K	ATRX(S598)
mTOR-PI3K	BCAS3(S838)
mTOR-PI3K	CTTN(S438)
mTOR-PI3K	EEF2K(S74)
mTOR-PI3K	FOXK2(T389+S392)
mTOR-PI3K	HDGFRP2(S454)
mTOR-PI3K	IBTK(Y996+S1004)
mTOR-PI3K	LARP1(S1056)
mTOR-PI3K	LARP4B(S568)
mTOR-PI3K	LIMA1(S225)
mTOR-PI3K	MAF1(S60+S68)
mTOR-PI3K	MKI67IP(S218+T240)
mTOR-PI3K	MREG(S213)

mTOR-PI3K	NDRG3(S331)
mTOR-PI3K	PCBP1(S264)
mTOR-PI3K	PLEKHA3(S244)
mTOR-PI3K	PXN(S137)
mTOR-PI3K	PXN(T136)
mTOR-PI3K	SPHKAP(T741)
mTOR-PI3K	USP24(S2561)
mTOR-PI3K	USP24(T2565)
p70S6K-PI3K	AKAP11(S18)
p70S6K-PI3K	CD2AP(S234)
p70S6K-PI3K	CPSF7(T67+Y72+S75)
p70S6K-PI3K	FKBP15(S1012)
p70S6K-PI3K	FKBP15(S346)
p70S6K-PI3K	PFKFB2(S483)
p70S6K-PKC	EPS8L2(S459)
p70S6K-PKC	EPS8L2(S466)
p70S6K-PKC	OR1A1(T280+Y288)
p70S6K-PKC	TPD52(S176)
p70S6K-PKC	UBL7(S230)
p70S6K-PKC	YAP1(S61)
PI3K-PKC	KIAA1598(S249)
PI3K-PKC	NCBP1(S22)
PI3K-PKC	PAK2(S197)
PI3K-PKC	SCNN1D(S6)
PKC-ROCK	EMD(S173)
Akt	ACACA(S29)
Akt	AFTPH(S411)
Akt	AHNAK(S4360)
Akt	AHNAK(T4766)
Akt	AKAP13(S790)
Akt	ANK3(S1326+S1327+T1334)
Akt	ANLN(S661)
Akt	ARAP1(S229)
Akt	ASXL2(S834+S842)
Akt	ATF2(S90)
Akt	ATXN2L(S559)
Akt	B4GALT1(S9+S11+S18)
Akt	BAG3(S377+S386)
Akt	BAG3(S385)
Akt	BAIAP2L1(T416)
Akt	BANP(S90)
Akt	BPTF(S1060+T1064)
Akt	BRD9(S588)
Akt	BZW2(S412+S414)
Akt	C17orf85(S25)
Akt	CAMSAP2(S611)
Akt	CAMSAP3(S1048)
Akt	CAMSAP3(S1048+S1051)
Akt	CBL(S452)
Akt	CCDC43(T139)
Akt	CCDC6(S254)
Akt	CCDC6(T436)
Akt	CHAMP1(S282)
Akt	CIC(S77)
Akt	CIZ1(T872)
Akt	CLIP1(S152)
Akt	CLNS1A(S197)
Akt	CP(T528)

Akt	CSTF2(S524)
Akt	CTNND1(T177)
Akt	CTPS1(S574+S575)
Akt	CYP2S1(T295+T304+T314)
Akt	DAP(S51)
Akt	DAXX(S688+S702)
Akt	DCAF5(S531)
Akt	DCP1A(S180)
Akt	DCTPP1(T170)
Akt	DIDO1(S898)
Akt	DIP2A(T279)
Akt	DIP2B(S153)
Akt	DNMT1(S394)
Akt	DPF2(Y172)
Akt	DSP(S2821+S2825)
Akt	DYNC1LI1(S207)
Akt	EIF4G1(S204)
Akt	EIF4G1(T211)
Akt	EIF5B(S135+S137)
Akt	ERBB2IP(Y920+S932)
Akt	EXOC1(S501)
Akt	EXOSC9(S392+S394)
Akt	FAM102B(S320)
Akt	FGFR1OP(S160)
Akt	FKBP15(S311)
Akt	FKBP3(S152)
Akt	FLNA(S2163)
Akt	FTH1(S183)
Akt	FUBP1(S630)
Akt	FYTTD1(S23)
Akt	GAB2(T287)
Akt	GAPDH(T75)
Akt	GAPVD1(S569)
Akt	GAPVD1(S758+S761)
Akt	GATAD2B(S129+S135)
Akt	GEMIN5(S778)
Akt	GOLGB1(S17)
Akt	GORASP2(S451)
Akt	GRIPAP1(S692)
Akt	GSG2(S179)
Akt	GTF2I(S679)
Akt	GYS1(S727)
Akt	HCFC1(S1507)
Akt	HMGCS1(S495)
Akt	HNRNPK(S116)
Akt	INF2(S1149)
Akt	KDM2A(S718)
Akt	KDM2A(T713+S718+T720)
Akt	KIAA0284(S655)
Akt	KIAA1429(S133)
Akt	KIAA1522(S161+T162)
Akt	KIAA1731(S407+S411+T415)
Akt	KLHL20(S263+T270+S273)
Akt	KRT19(S10)
Akt	LAD1(S121+S123)
Akt	LARP4(S597)
Akt	LRIG1(S1044)
Akt	MACF1(S3927)

Akt	MAP4(S507+T521)
Akt	MCM3(S711)
Akt	MCRS1(S282)
Akt	MFF(T138)
Akt	MIIP(S303)
Akt	MKI67(S2471)
Akt	MKI67IP(T227)
Akt	MKI67IP(T234+T238)
Akt	MTMR3(S613)
Akt	MTMR4(S610)
Akt	MUC5AC(T2223+S2224+T2225)
Akt	MYL9(S20)
Akt	MYL9(T19+S20)
Akt	NACA(S166)
Akt	NAGK(S76)
Akt	NAP1L1(T39)
Akt	NCAPG(S1015)
Akt	NCOR2(S2057+S2065+S2068)
Akt	NRP1(S894)
Akt	NUCKS1(S19)
Akt	NUMA1(S1945)
Akt	NUMA1(T2106)
Akt	NUP153(S343)
Akt	OR6C65(T277+S278+Y288)
Akt	PAK2(S141)
Akt	PBRM1(S39)
Akt	PDIA6(S428)
Akt	PHF8(S857)
Akt	PHKA1(S868+T877+S884)
Akt	PHKA2(S1015)
Akt	PKN1(T774)
Akt	PKP3(S313+S314)
Akt	POGZ(S333)
Akt	PPP1R2(S87)
Akt	PREX1(S319)
Akt	PTGES3(S151)
Akt	PXN(S340)
Akt	RAB11FIP1(S501)
Akt	RAB11FIP5(S176)
Akt	RAB3IP(S288)
Akt	RABGEF1(Y601)
Akt	RABL6(S425+S427)
Akt	RAD21(S449)
Akt	RAD50(T690)
Akt	RAD9A(S277)
Akt	RAI1(S683)
Akt	RAI1(S924)
Akt	RALGAPA1(S773)
Akt	RALGAPA1(S797)
Akt	RANBP2(S2263+S2270)
Akt	RBBP6(S1277)
Akt	RFFL(S240)
Akt	RHBDL1(S193+S202+Y214)
Akt	RPL24(T83)
Akt	RPLP0(S304+S307)
Akt	RPLP0(S307)
Akt	RPRD2(S976)
Akt	RPS17L(S113)

Akt	RPS17L(S115)
Akt	RPS6KA1(S221)
Akt	RWDD2B(T175)
Akt	SAC3D1(S402)
Akt	SAP30(S131+S138)
Akt	SAP30BP(S52)
Akt	SCAF1(S239)
Akt	SCAF1(S453)
Akt	SCAF1(T994+S997)
Akt	SCFD1(S373)
Akt	SCNM1(S183)
Akt	SCRIB(S1348)
Akt	SEC16A(S414+T415+S417)
Akt	SF1(S80+S82)
Akt	SF3B1(S129+T142)
Akt	SF3B2(T780)
Akt	SGSM1(S67)
Akt	SH2D3A(S125)
Akt	SIPA1L1(S174)
Akt	SIPA1L1(Y206+T209)
Akt	SLC3A2(S134)
Akt	SMARCA5(S66)
Akt	SMEK1(S771)
Akt	SMEK2(S840)
Akt	SNAPIN(S133)
Akt	SON(S1556)
Akt	SOX6(T180+S183)
Akt	SRRM2(S1401+S1404)
Akt	SSR3(S11)
Akt	STK24(T184)
Akt	STRN3(S257)
Akt	STRN4(S276)
Akt	SYAP1(S267)
Akt	TAB2(S524)
Akt	TBC1D10B(S656+S661)
Akt	TBC1D22A(S132)
Akt	TCF3(S134+S139)
Akt	TDRD7(S859)
Akt	TEX15(S194+S195+Y210)
Akt	TFDP1(S23)
Akt	TNS3(S776)
Akt	TP53BP1(S294)
Akt	TSC2(S1798)
Akt	TWF1(S143)
Akt	U2AF2(T119+T124+S142)
Akt	UBAP2L(T425)
Akt	UNG(T60)
Akt	UPF1(S1107)
Akt	USP13(T122)
Akt	USP32(S1423)
Akt	USP5(T623)
Akt	VAMP4(S30)
Akt	VTI1B(S138)
Akt	WDHD1(S367)
Akt	YTHDF1(S291)
Akt	ZFC3H1(T823)
Akt	ZNF276(S599)
Akt	ZNHIT3(S80)

Akt	ZNRF1(S52+S53)
Akt	ZZEF1(S1518)
CAMK2	ARHGEF2(S151)
CAMK2	ARSI(Y263)
CAMK2	CAC1A(S2275+T2277)
CAMK2	CCDC86(S18+S21)
CAMK2	DNMT3A(S6+S7+T12)
CAMK2	HUWE1(S649)
CAMK2	JUND(S255+S259)
CAMK2	KCTD15(S35+S38)
CAMK2	MKI67IP(S230+T234+T238)
CAMK2	PPP1R11(T109)
CAMK2	PRKCD(S664)
CAMK2	PTGES3(S113)
CAMK2	RPRD2(S628)
CAMK2	RRM2(S377)
CAMK2	SCRIB(S1439)
CAMK2	SMEK1(S117+S126)
CAMK2	TRIM33(T1102+S1105)
EGFR	AHNAK(S5448)
EGFR	ARHGAP17(S625)
EGFR	EHMT2(S140)
EGFR	EIF4G1(S1092)
EGFR	FASN(T2204)
EGFR	GCFC1(S557+S558)
EGFR	GORASP2(T433)
EGFR	ITGB4(S1457)
EGFR	LSM14A(S178+S182)
EGFR	NSFL1C(S114)
EGFR	PDS5B(S1177)
EGFR	RIPK2(S363)
EGFR	SRRM1(S769+S775)
EGFR	SRRM2(S2118+S2121)
EGFR	TNRC6B(S1816)
EGFR	TOP2A(S1393)
EGFR	WNK2(S1862)
ERK	CEP128(S217)
MEK	AHCTF1(S1283)
MEK	BCL2L11(S77)
MEK	ERF(S327)
MEK	MAPK1(Y187)
MEK	MAPK3(Y204)
mTOR	ALKBH5(S361)
mTOR	ASAP2(S822)
mTOR	CASZ1(S741)
mTOR	DOS(S361)
mTOR	EIF4EBP1(S65)
mTOR	EIF4EBP1(T70)
mTOR	EIF4EBP1(Y34+T45)
mTOR	EIF4EBP2(T46)
mTOR	FA13B(S14)
mTOR	FOXK1(S441+S445)
mTOR	FOXK2(S428)
mTOR	HN1(S87+S88)
mTOR	KLF16(S224+S246)
mTOR	MKI67(S2223)
mTOR	MKI67(S2708)
mTOR	NDRG3(S333+S334)

mTOR	PHF2(S539)
mTOR	RPS10P5(S157)
mTOR	ULK1(S556)
mTOR	ULK1(S638)
mTOR	WIZ(S1146)
mTOR	ZFP106(S861)
p70S6K	CELSR1(S1320)
p70S6K	CIB3(T6+T9+Y16)
p70S6K	CLSPN(S1289)
p70S6K	ITLN2(S18)
p70S6K	KRT18(S34)
p70S6K	MAP6(S793)
p70S6K	MAP7(S365)
p70S6K	MED1(S953)
p70S6K	PPP1R10(S591)
p70S6K	PRKAG1(Y280+T284+T287)
p70S6K	RFX7(S1081)
p70S6K	RHBDF1(S51)
p70S6K	THRAP3(S682)
p70S6K	WBP11(S237)
p70S6K	WDHD1(S333)
p70S6K	ZNF217(S795)
PI3K	AHNAK(S570)
PI3K	DLG5(S1075)
PI3K	DOCK7(S2131)
PI3K	EHBP1L1(S310)
PI3K	FAM102B(S350)
PI3K	FAM21C(S288)
PI3K	FBP2(T13)
PI3K	FKBP15(S344+S356)
PI3K	FOXK2(S369+S373)
PI3K	GGT5(S474)
PI3K	KLC1(S524)
PI3K	LARP1(S824)
PI3K	MUC5B(T490)
PI3K	PACS1(S355)
PI3K	PAK2(S141+T154)
PI3K	PKP3(S313)
PI3K	PREX1(S1179)
PI3K	PREX1(S1182)
PI3K	PUM1(S75)
PI3K	SPAG9(S203+T217)
PI3K	STMN1(S16)
PI3K	ULK1(S623)
PKC	ABCF1(S140)
PKC	AHNAK(S5841)
PKC	ARFGAP1(T135)
PKC	CCDC88C(S1887)
PKC	CDC40(S45)
PKC	EIF4B(S442+S445)
PKC	EIF4B(S459)
PKC	GOLGA5(S116)
PKC	MAP2K2(S26)
PKC	MARK3(S419)
PKC	MARK3(T530)
PKC	MKI67(S713)
PKC	NAA10(S182)
PKC	NAV1(S17)

PKC	NFATC2IP(S366)
PKC	NOP58(S502+S514)
PKC	NPM1(S70)
PKC	NRBF2(S120)
PKC	NUP214(T1021)
PKC	OR5K4(T134)
PKC	PPP4R2(S224)
PKC	PVRL1(S422)
PKC	RANBP2(S1160)
PKC	SHANK1(S958)
PKC	SPTBN2(S2384)
PKC	TPT1(S46)
PKC	TRA2B(T201)
PKC	TWISTNB(S328)
PKC	YTHDC1(S424)
ROCK	AAK1(S637)
ROCK	AHNAK(S819)
ROCK	DIP2A(S94)
ROCK	DIP2B(S100)
ROCK	DIP2B(Y98)
ROCK	FBXW12(S286)
ROCK	IQSEC1(S180)
ROCK	MKI67(S1329)
ROCK	MKI67(S1937)
ROCK	MKI67(T1315)
ROCK	MPRIP(S1014)
ROCK	PPP1R12A(S445)
ROCK	PPP1R12A(S903)
ROCK	PPP1R13L(S134)
ROCK	SAPCD2(T219)
ROCK	SPECC1L(S887)
ROCK	SPTBN1(S2341)
ROCK	SVIL(T852)
ROCK	SVIL(Y850)
ROCK	WHSC2(S363)

Appendix 2

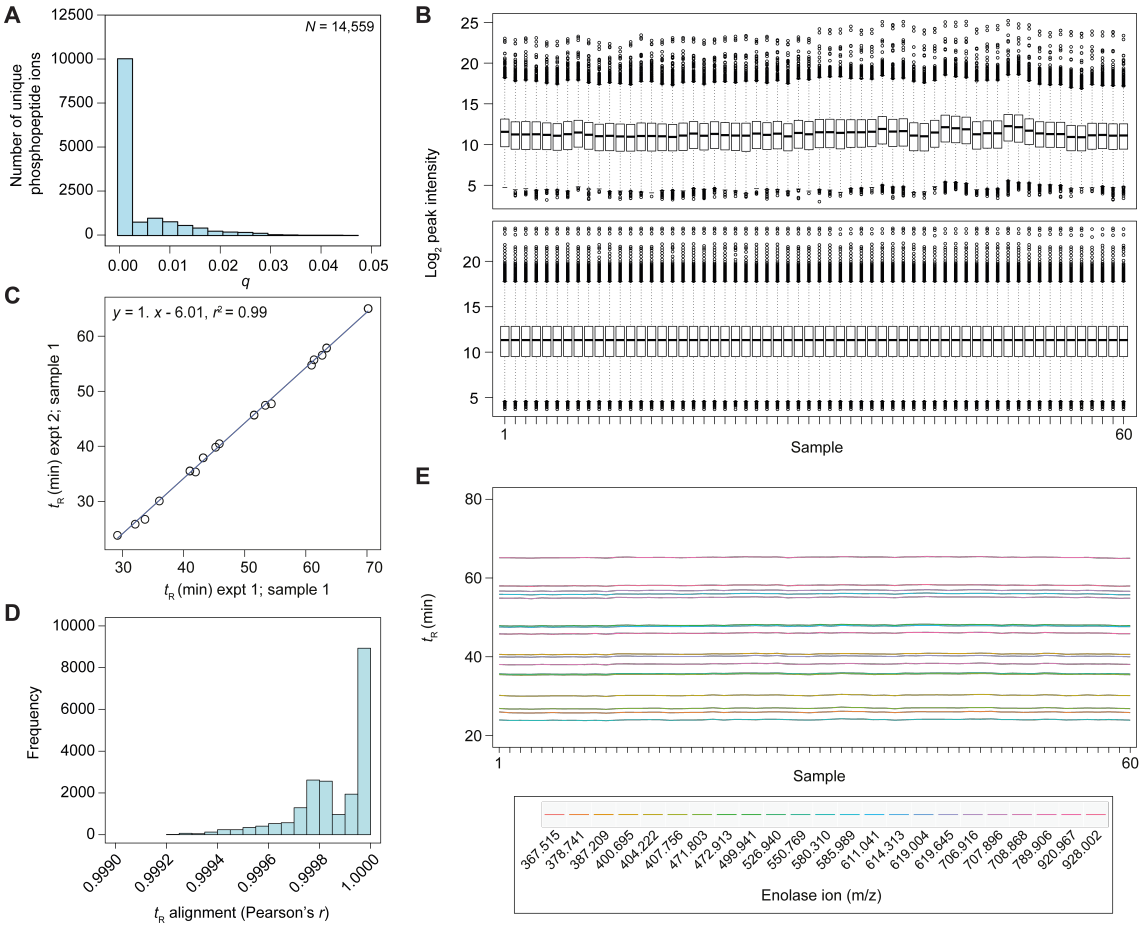
Summary of CTAM groups (Chapter 4).

CTAM group	**Number of members (*t*)**
Akt	209
Akt-CAMK2	39
Akt-CAMK2-EGFR	15
Akt-CAMK2-EGFR-ERK	6
Akt-CAMK2-EGFR-ERK-MEK	4
Akt-CAMK2-EGFR-ERK-MEK-mTOR	4
Akt-CAMK2-EGFR-ERK-ROCK	1
Akt-CAMK2-EGFR-mTOR	1
Akt-CAMK2-EGFR-p70S6K	1
Akt-CAMK2-ERK	2
Akt-CAMK2-ERK-MEK	1
Akt-CAMK2-ERK-MEK-mTOR	1
Akt-CAMK2-mTOR-p70S6K-PI3K	1
Akt-CAMK2-mTOR-PI3K	1
Akt-EGFR	9
Akt-EGFR-ERK	4

Akt-EGFR-ERK-mTOR	1
Akt-EGFR-MEK	1
Akt-ERK	1
Akt-ERK-MEK	1
Akt-ERK-mTOR-p70S6K-PI3K	1
Akt-mTOR	5
Akt-mTOR-p70S6K-PI3K	34
Akt-mTOR-p70S6K-PI3K-PKC	1
Akt-mTOR-PI3K	45
Akt-p70S6K	2
Akt-PI3K	6
Akt-PKC	1
CAMK2	17
CAMK2-EGFR	4
EGFR	17
EGFR-ERK	1
EGFR-ERK-MEK	1
EGFR-MEK	3
EGFR-p70S6K	1
EGFR-PI3K-PKC	1
ERK	1
ERK-mTOR	1
ERK-p70S6K	1
MEK	5
MEK-PKC	1
mTOR	22
mTOR-p70S6K	1
mTOR-p70S6K-PI3K	7
mTOR-PI3K	22
p70S6K	16
p70S6K-PI3K	6
p70S6K-PI3K-PKC	1
p70S6K-PI3K-PKC-ROCK	1
p70S6K-PKC	6
PI3K	22
PI3K-PKC	4
PKC	29
PKC-ROCK	1
ROCK	20

Appendix 3

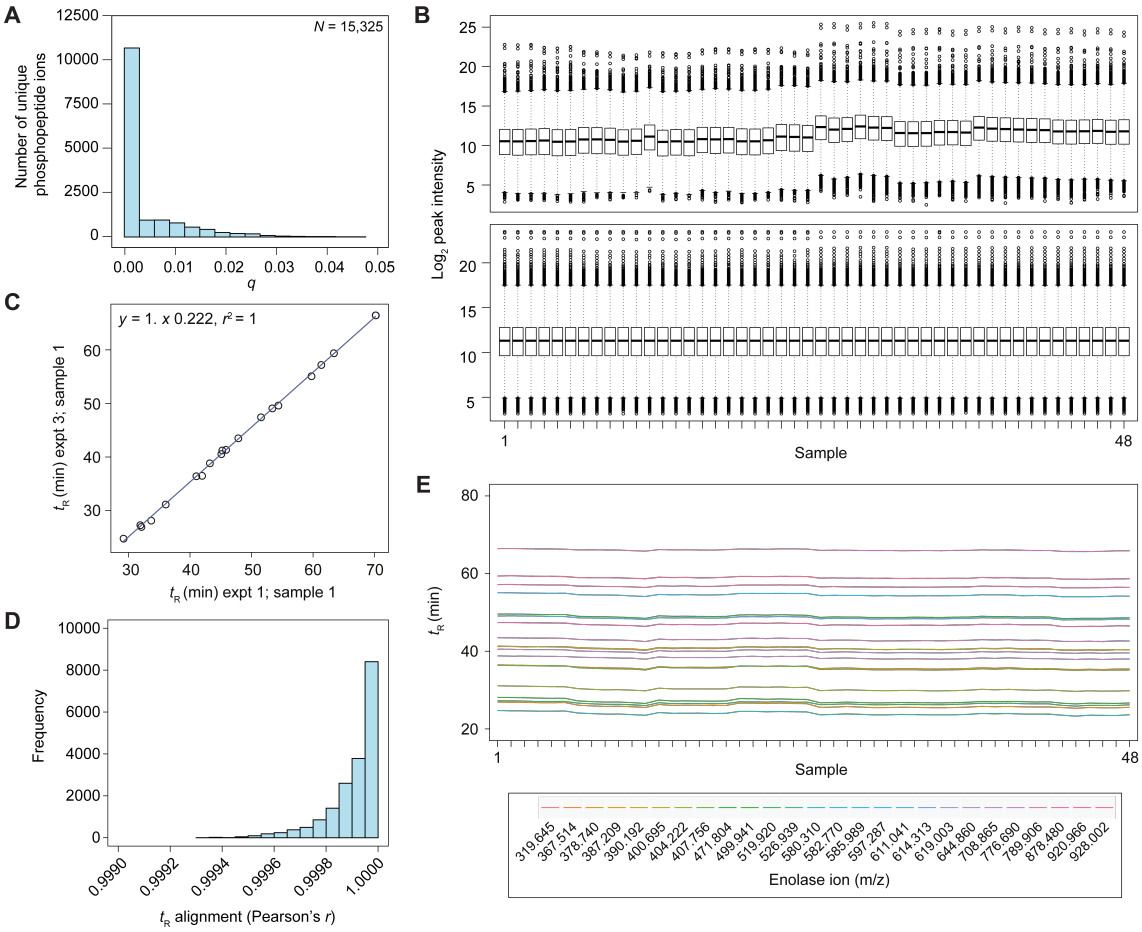
Quality control analysis for EGF and IGF-1 stimulation experiments (Chapter 5).



Appendix 3. (A) Distribution of *q*-values for the identified phosphopeptides. (B) Representative pairwise alignment of enolase ion retention times used to predict phosphopeptide retention times in samples where the phosphopeptide was not identified. (C) Distribution of Pearson's *r* values for each pairwise enolase ion retention alignment. (D) Upper panel: box plots demonstrating the distribution of $\log_2$-transformed phosphopeptide peak intensities pre-normalisation. Lower panel: $\log_2$ phosphopeptide peak intensities distributions post-quantile normalisation. (E) Enolase ion retention time vs sample (in chronological order of nLC-MS/MS run). Each line represents an individual enolase ion.

Appendix 4

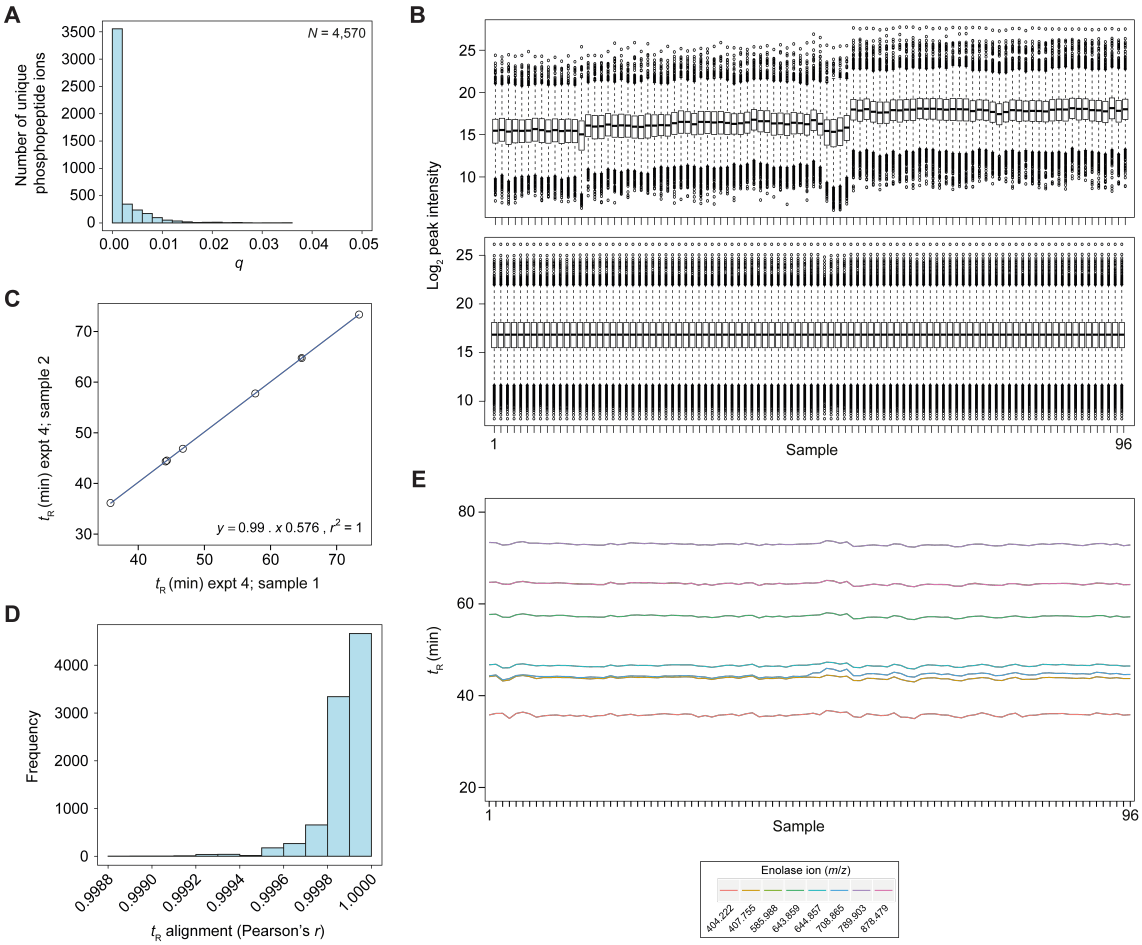
Quality control analysis of resistant cell-line phosphoproteome comparisons (Chapter 6).



Appendix 4. (A) Distribution of *q*-values for the identified phosphopeptides. (B) Representative pairwise alignment of enolase ion retention times used to predict phosphopeptide retention times in samples where the phosphopeptide was not identified. (C) Distribution of Pearson's *r* values for each pairwise enolase ion retention alignment. (D) Upper panel: box plots demonstrating the distribution of $\log_2$-transformed phosphopeptide peak intensities pre-normalisation. Lower panel: $\log_2$ phosphopeptide peak intensities distributions post-quantile normalisation. (E) Enolase ion retention time vs sample (in chronological order of nLC-MS/MS run). Each line represents an individual enolase ion.

Appendix 5

Quality control analysis of MCF-G and -P cell-line EGF stimulation experiments (Chapter 6).



Appendix 5. (A) Distribution of q-values for the identified phosphopeptides. (B) Representative pairwise alignment of enolase ion retention times used to predict phosphopeptide retention times in samples where the phosphopeptide was not identified. (C) Distribution of Pearson's r values for each pairwise enolase ion retention alignment. (D) Upper panel: box plots demonstrating the distribution of $\log_2$-transformed phosphopeptide peak intensities pre-normalisation. Lower panel: $\log_2$ phosphopeptide peak intensities distributions post-quantile normalisation. (E) Enolase ion retention time vs sample (in chronological order of nLC-MS/MS run). Each line represents an individual enolase ion.